EXPERIMENTAL PSYCHOLOGY:
An Introduction

Under the Editorship
of
GARDNER MURPHY

EXPERIMENTAL PSYCHOLOGY:

AN INTRODUCTION

By

LEO POSTMAN
University of California

and

JAMES P. EGAN
University of Indiana

KALYANI PUBLISHERS
LUDHIANA

10

87 = 82

EXPERIMENTAL PSYCHOLOGY : An Introduction

Price : Rs. 19.80

*Reprinted in India by arrangement with Harper & Row,
Publishers, Incorporated*

Published by Usha Raj Kumar for Kalyani Publishers,
Ludhiana and Printed by Mohan Makhijani at Rekha
Printers Pvt. Ltd., New Delhi-110020

To
Dorothy
and
Elizabeth

CONTENTS

LIST OF EXPERIMENTS

PREFACE

THIS text was written for the student who has some knowledge of general psychology and who now needs to acquaint himself with experimental methods and laboratory procedures. Experimental methods and laboratory procedures cannot, of course, be divorced from psychological facts and principles. Our goal was, therefore, to give a survey of the main empirical findings and functional relationships in selected areas of experimental psychology with special emphasis on the control, manipulation, and measurement of variables.

In each of the areas selected for treatment we have presented a few experimental exercises. We are fully aware of the necessity to organize experimental courses around the facilities available in a given laboratory. We are also aware of the differences in tradition and interests which make for diversity in experimental courses—a diversity that has proved so fruitful and stimulating to research. Probably the time has not yet come to standardize our laboratory courses. We offer our exercises, therefore, not primarily in the spirit of a laboratory manual although the instructor can use these experiments in his laboratory. These exercises are intended more as concrete illustrations which will supplement the text in their emphasis upon specific methods of collecting and interpreting experimental data.

Although we call this book *Experimental Psychology*, we do not for a moment think that we have come anywhere near treating all the different areas and problems that experimental psychology now claims. The horizons of experimental psychology are now so broad that the best an introductory course may hope to offer is a representative sample of methods and results. Such a selection of topics represents one way of introducing the student to this wide field.

We have not attempted to document the text extensively with references to the original sources. We feel that the beginner in the field might be unduly burdened by the large number of references

that would be required for adequate documentation. There are, moreover, many facts and principles which have been fully established and accepted into the body of psychological knowledge and which need no longer be identified with any single source. We have, therefore, adopted the practice of most of the elementary textbooks in psychology and have confined ourselves principally to selected references at the end of each chapter.

In writing this text we have at times leaned heavily upon the more advanced treatises and handbooks. We shall not try to enumerate these sources. Our indebtedness to specific men and their works will be obvious.

We are grateful to the editor of this series, Professor Gardner Murphy, for his help. and encouragement. We are indebted to William F. John who labored patiently over many of the illustrations. Elizabeth Egan kindly helped in the preparation of the manuscript. Lorraine Lerman helped to read the proof. Our most faithful and patient helper in the preparation of the manuscript was Dorothy L. Postman. Without her efforts, it would have been difficult indeed to complete this book.

Many publishers and authors have kindly consented to the reproduction of original material. We are greatly indebted to them for their courtesy. Specific acknowledgments appear in the text.

May, 1949 L. P.
Cambridge, Massachusetts J. P. E.

EXPERIMENTAL PSYCHOLOGY:
An Introduction

»»» 1 «««

THE SCOPE OF EXPERIMENTAL PSYCHOLOGY

Experimental Psychology as Method. The field of experimental psychology is not defined in terms of subject matter alone. Many disciplines are concerned with what people do and why they do it. It is the particular way in which experimental psychology proceeds to study behavior that characterizes it. Experimental psychology attempts to apply the rules of scientific method to its subject matter: to discover the lawful relationships that govern behavior. Whether the behavior be as simple as pressing a telegraph key or as complex as painting a picture, experimental psychologists search for the determinants of these actions in the behaving organism and its environment.

Variables. Whatever behavior we study, we usually ask what its antecedents are, i.e., under what conditions it occurs. We aim to understand the determinants of behavior by discovering the *relations* among clearly defined *variables*. In its broadest sense, a variable is a characteristic or attribute that can take on a number of values. Thus, the number of items that an individual solves on a particular test is a variable. The speed with which we respond to a signal, the readiness with which we are influenced by propaganda, the size of our pupil in different degrees of illumination, these and many others are examples of variables which experimental psychology studies.

The experimenter distinguishes between two basic types of variables: *independent* and *dependent*. The phenomena which we wish to explain and predict are the *dependent variables*. These variables are called dependent because they depend upon the occurrence of particular antecedent conditions. In experimental inquiry we manipulate the antecedent conditions in order to discover the ways in which they determine the dependent variables. The antecedent conditions that the experimenter manipulates freely are called the

independent variables. In the examples given above, the number of items solved by a subject is the dependent variable; the nature of the test and the conditions of its administration may be the independent variables. Similarly, we may choose intensity of illumination as our independent variable, manipulating it freely, and measure the pupillary area of the eye as the dependent variable. In all such experiments our ultimate aim is to discover a principle or law relating the dependent to the independent variables over a wide range of values.

Stimuli and Responses as Variables. The variables of experimental psychology are stimuli and responses: events in the environment and the organism's responses to them.

Any stimulus is a *selected aspect* of the total environmental situation, and any response is a selected characteristic of the total complex of behavior occurring at any moment. Having selected such aspects of the environment and of behavior for study, we are then faced with the problem that stimuli and responses are not fixed things or events which can be reproduced identically from one moment to another. Let us illustrate with the aid of a common experimental situation.

Suppose we require a subject to learn a series of words. These words are printed on cards and are repeatedly exposed to the subject. Upon successive presentations of the same word, the position of the card with respect to the subject's eyes may vary from trial to trial. Yet, we do not distinguish among the different presentations of the same card but simply count the number of times the stimulus card was presented. Clearly, then, what is treated as *the* stimulus in this experiment is actually a selected *class of events* having certain properties in common. In this case, the common property is the ability of the subject to read the word when the card is presented to him. In short, the concept of the stimulus is a *generic* one.

What we have just said with respect to the generic nature of stimuli applies with equal force to the concept of response. In studying the process of learning telegraphic code, for example, the subject's response is the tapping out of a word or phrase. We consider all responses as equivalent that have a certain effect: tapping out a given word. Though equivalent in effect, these responses may, and often do, differ as regards specific movements. The word *the* may

involve slightly different actions of the fingers each time it is sent. Thus, when we count the responses, we really count the occurrence of a certain *type of response.* Before we can count responses, however, we must first decide over what range the responses may vary and still be included under a given type. As long as the letters tapped out are *t-h-e,* we accept them as the response *the.* Of course, if we were interested in the action of particular muscle groups and not the number of words sent per minute, we would define response in a very different way. Responses, like stimuli, are selected classes of events that have certain critical properties in common.

Experimental Control of Variables. Suppose we wish to study the relation between a dependent variable, y, and an independent variable, x: the number of trials required to learn a list of words as determined by the length of the list. Clearly, y, the number of trials, depends not only on the length of the list, x, but on a number of other factors, z, u, w, such as the difficulty of the words, the speed of their presentation to the subject, and the time interval between successive trials. What we are interested in, however, is the specific relation between y and x, the number of trials to learn and the length of the list. If we let z, u, and w vary in an uncontrolled way, we cannot attribute changes in y to changes in x. We can tease out the relation between y and x best by holding z, u, and w at constant values while varying x. In our example, we would prepare lists of different length (variable x) but of equal difficulty (variable z), and present these lists at the same rate (variable u), and with a constant time interval between successive trials (variable w). Thus, we would determine variations in y as a function of variations in x with z, u, and w held constant.

The specific relation that we obtain between y and x depends upon the particular constant values that we assigned to z, u, and w. We can repeat the experiment using different values for z, u, and w, and obtain a new functional relation between y and x. The variables z, u, and w, which are constant during a particular experiment but which may be varied from experiment to experiment, are called *parameters.*

But the reader may object that we do not know all of the variables which might influence our dependent variable, y. Will holding z, u, and w constant be sufficient if the unknown determinants, m, n, p,

etc., are allowed to vary in an uncontrolled manner during the experiment? What if eye movements and changes in room illumination are uncontrolled? Perhaps these are important determinants of y. We must hope that such uncontrolled factors vary "at random," that they do not exert a *systematic* effect on the results of the experiment. Variations in eye movements occur during the learning of *each* list. We trust that the magnitude of illumination changes will not be greater for one length of list than for another. In short, we assume that the effects of the uncontrolled variables, m, n, and p, will, *on the average*, be the *same* for all the conditions of the experiment. Then, the relation between y and x will not be systematically affected by m, n, p, etc.

The use of several subjects for each value of the independent variable, x, serves to minimize the effects of individual differences. If we used a single subject for each length of list, we would not know whether to ascribe changes in the dependent variable, y, to changes in the independent variable, x, or to individual differences among the subjects. By using several different subjects under each condition, we hope to average out the effects of such individual differences. Another way to do so is, of course, to have each subject learn each length of list.

Experimental and Control Groups. In studying the relation between y and x, it is often necessary first to make sure whether or not y depends upon x at all. If, for example, we are interested in establishing whether or not vitamin A has an effect on night vision, we may conveniently answer this question by using an *experimental group* and a *control group*. The same tests of vision are given to both groups. We select the groups so that they will be as much alike as is practically possible, for example, in age, health, diet, general visual functions, and night vision in particular. The essential difference between the two groups is that the experimental group receives a certain dose of vitamin A, whereas the control group does not. We compare the performance of the experimental group under conditions of night vision with that of the control group before, during, and after administration of vitamin A. If the experimental group shows significantly superior performance over the control group, we ascribe this difference to the effects of vitamin A. We had to use a control group because we did not know what

changes in night vision might occur independently of vitamin A, say, as the result of practice, increased motivation as the experiment progresses, etc. The control group provides us with an empirical base line against which we can evaluate the effects of the critical variable. If the effectiveness of the variable is established, we may wish to quantify the relation between the amount of x (dose of vitamin A) and the changes in y (variations in night vision).

Statistical Significance of Experimental Results. Even in a well-designed experiment the uncontrolled factors will not be completely averaged out. Such uncontrolled factors may be responsible for observed changes in the dependent variable, y. The experimenter must decide whether the changes he has observed in y are due to his independent variable, x, or to uncontrolled variables, m, n, p, etc. It is at this point that statistical techniques play an important role. By such techniques we may estimate how probable it is that the observed changes in y are due merely to the random action of uncontrolled variables. If that probability is very low, say, one in a hundred, we conclude that it is the independent variable, x, which is responsible for the change in y. We never prove with complete certainty that one variable influences another. We can be only more or less certain that the relation between y and x is not due to chance. This degree of certainty is sometimes so high that we speak of positive proof.

Description and Explanation. Experimental inquiry necessarily begins with a *description* of the phenomena under investigation. In our daily lives we continually describe our own actions and those of others, and we have developed a rich variety of terms for doing so. In our urge to understand ourselves and others, we often fail to distinguish between description and explanation. We explain away many phenomena of perception by describing the attributes of our experiences and then endowing the physical objects with these attributes. A sheet of white paper in shadow looks white, and coal in sunlight looks black because "this paper is white and coal is black." Clearly, such a description of our experiences does not explain perception.

As psychologists we need to make the distinction between description and explanation with utmost care. Explanation comes with the establishment of lawful functional relationships between dependent

and independent variables. In our example, the description of paper as white and coal as black would be the dependent variable determined by a host of stimulus conditions and properties of the organism.

Forms of Behavior Studied in Experimental Psychology. The forms of behavior that can be observed are infinitely varied. From this variety we must select some forms rather than others for experimental investigation. There are no cut-and-dried rules for such selection. Some psychologists study the specific movements of organisms, and others put emphasis on acts and their outcomes, while again others concentrate upon the interaction of individuals with their social environment. The study of all these forms of behavior has a legitimate place in experimental psychology.

We may illustrate what is meant by different forms of behavior with the aid of the following examples. Let us begin with a relatively simple and circumscribed response. When a puff of air is directed at the cornea of the eye, the human subject responds by blinking. This response is highly reliable since it will occur almost invariably a very short time after the puff strikes the eye. This response is common to virtually all people and is part of the innate equipment of the organism. The dependent variables which we may investigate in connection with this response are the time interval between the air puff and the blink, the form of the response, and its amplitude. We can readily quantify these variables.

In studying the eyeblink, we may be interested in the particular form of the movement. In many experimental situations, however, we are not especially concerned with the specific nature of movements. Rather we want to know whether or not a response occurs under given conditions, and how often it occurs over a certain period of time. Thus, referring to our example of learning code, we usually count the number of words that an operator can send per minute. The dependent variable in this type of investigation is a simple counting variable: number of words. We do not usually record the particular movements which the operator uses, but study only the effects of these movements. We do not mean to imply that specific movements are unimportant. Clearly, a pressure on the telegraph key that is too weak or too violent detracts from the efficiency of the operator. However, as long as the operator improves

with practice, we may not be concerned with the exact way in which he depresses the key.

Some problems in which experimental psychologists are interested are broader than particular movements or their effects. The social values held by an individual may enter as a variable. It is well known, for example, that the effectiveness of propaganda and suggestion may vary with their source. In experimental investigations of these phenomena, we can systematically vary the sources of propaganda and suggestion and measure the resulting changes in acceptance and belief. We may find that it is the "prestige" of the propagandist which is an important determiner of the success of propaganda. The prestige which one individual holds for another does not necessarily depend upon a particular kind of act or a specific set of movements. A variety of acts and a variety of movements may create equal degrees of prestige. It is prestige, independent of the particular acts and movements from which it originated, which is often the relevant variable for the experimental social psychologist.

Experience and Behavior. Among the data of experimental psychology are not only movements and acts but also statements that subjects make in reference to their experience. Are we justified in using such "subjective" data? If the conditions under which the subject reports on his experience are clearly and rigorously defined, such statements are legitimate data for psychology.

Much experimentation in the sensory processes and perception must rely on the subject's reports obtained under fully specified physical conditions. We stimulate a subject's eye with a particular wave length of light, and he pronounces the word "red." We change the wave length to a new value, and the subject now says "green." We repeat such trials over and over with different wave lengths and the subject uses the words that designate colors with great consistency. Furthermore, most subjects agree with each other in their use of color names in response to different wave lengths. The uniformity of the statements which the subjects make when their eyes are stimulated by various wave lengths has led psychologists to infer that there is the *dimension of experience:* hue. In a similar way, brilliance and saturation have been established as dimensions of visual experience. In all sense modalities, experience has many

modes of variation, and these dimensions are known from the consistency of the verbal reports given in response to stimulation. For the psychologist, statements about experience are useful only if the conditions under which these statements are made can be accurately described.

The study of experience includes more than the investigation of sensory attributes. Sensory experience is organized: we see things rather than their attributes, books and tables rather than redness and squareness. To find the principles of perceptual organization is one of the major endeavors of present-day experimental psychology, which cannot be easily carried out without benefit of the subject's reports on his world.

Of course, verbal report provides us with information other than the attributes or organization of experience. The responses made in a free-association experiment or a subject's reports in problem-solving situations are examples of other types of verbal behavior.

No matter what form of behavior we investigate, our main problem is to define the independent and dependent variables, to investigate the functional relations between them, and wherever possible to express our results in quantitative terms.

General References

Andrews, T. G. (ed.), *Methods of psychology*, New York: John Wiley and Sons, 1948, Chap. 1.

Cohen, M. R., and Nagel, E., *An introduction to logic and scientific method*, New York: Harcourt, Brace and Co., 1934, Part II.

Fisher, R. A., *The design of experiments*, Edinburgh: Oliver and Boyd, 1937.

Murphy, G., Murphy, L. B., and Newcomb, T. M., *Experimental social psychology*, New York: Harper & Brothers, 1937, Chap. 1.

Peters, C. C., and Van Voorhis, W. R., *Statistical procedures and their mathematical bases*, New York: McGraw-Hill Book Co., 1940, Chap. 16.

Pratt, C. C., *The logic of modern psychology*, New York: The Macmillan Co., 1939.

Skinner, B. F., The generic nature of the concepts of stimulus and response, *J. Gen. Psychol.*, 1935, *12*:40-65.

Stevens, S. S., Psychology and the science of science, *Psychol. Bull.*, 1939, *36*:221-263.

»»» 2 «««

THE PSYCHOPHYSICAL METHODS

ONE of the oldest problems in psychology is the relation between variations in physical stimulation and reported experience. Historically, this field of investigation has often been tied to such philosophical issues as the mind-body problem and the nature and meaning of "subjective experience." Today psychologists are content to leave the solution of philosophical problems to the philosopher, but there still remains a large area of research of great theoretical as well as practical importance: What are the lawful relationships between the measurable characteristics of the stimulus, on the one hand, and the reportable attributes of sensory experience, on the other? It is to this question that the division of experimental psychology known as *psychophysics* addresses itself. Psychophysics is indeed the earliest branch of experimental psychology. The theoretical value of psychophysics lies in the fact that it provides one important experimental approach to the study of the sensory processes and of judgment. As for its practical value, the knowledge gained by the methods of psychophysics has received increasingly wide application in such fields as personnel selection and equipment design.

In subsequent chapters we shall present the methods and results of psychophysics as they apply to the special senses—vision, hearing, the chemical senses, and the cutaneous senses. In the sections that follow we shall be concerned with the procedures and quantitative methods of psychophysics in so far as they are general methods applicable to a variety of specific problems.

The Basic Problems of Psychophysics

In studying the relation between the characteristics of the stimulus and the attributes of experience, certain specific experimental questions are asked.

Detection of minimal stimuli. What is the minimum of stimulation required for the detection of a stimulus? What kind of stimulus is needed? How intense must it be in order that a subject may reliably distinguish between its presence and absence? Obviously, the minimum amount of stimulation required will vary with the conditions of testing. To be barely detectable, a tone has to be less intense in a sound-treated room than in a noisy one, and, similarly, a weaker light is needed in a dark room than in a well-lit one. But for each condition of testing and for each subject, such a minimum value of a given stimulus can be estimated.

Detection of minimal stimulus differences. What is the minimal difference, qualitative or quantitative, needed between two stimuli so that they can be reliably recognized as different by a subject? For example, how different do two light stimuli have to be in wave length for a detectable difference in hue? How great a difference in intensity of light is required for a discrimination of brilliance? Again, the minimum value of the difference will vary from one testing situation to another and from one subject to another.

Judgment of relations among stimuli. The experimental problems of psychophysics are not limited to the study of stimuli and stimulus differences that are barely detectable. The judgment of stimuli well above the minimum needed for discrimination defines another important area of investigation. Under what conditions, for example, are two stimuli judged to be equal or as standing in a certain relation to one another? What is the extent of error when subjects attempt to equate two stimuli with respect to quality or quantity? How reliably can subjects respond to a stimulus as being half as intense or twice as intense as another stimulus? These are just a few illustrations of the problems which arise in connection with judgment of relations among stimuli.

The Basic Concepts of Psychophysics

Sensitivity. The organism is equipped with a number of receptor organs specialized to respond to particular energy changes in the environment. The receptors of the eye are responsive to light within a certain range of wave lengths, the receptors of the ear to sound waves within a certain range of frequencies, and so on. The action of these receptor organs constitutes an important link in the chain

of responses which occurs between the application of the stimulus and the subject's response. The capacity of the receptor organs and other reaction systems in the organism to respond selectively and differentially to physical stimulation we designate as *sensitivity*. The laws governing sensitivity are inferred, with the aid of psychophysical procedures, from the variations in response resulting from variations in stimulation. Our experimental measurements allow us to distinguish two types of sensitivity: *absolute* and *differential*.

Absolute sensitivity defines the limits of the organism's capacity to respond to stimulation. It is inversely related to the minimum stimulus which can be detected reliably by a subject. *Differential sensitivity* defines the organism's capacity to respond to differences, both qualitative and quantitative, between stimuli. It is inversely related to the minimum difference between stimuli needed for reliable discrimination.

Thresholds. Some stimuli are so weak that they always fail to evoke an effective response in the organism; others are so intense that they never fail to produce a reaction. The line separating these two kinds of stimuli—those never yielding responses and those always yielding responses—can never be sharply drawn; rather, the transition from one to the other is gradual and continuous.

Suppose we wish to measure a subject's absolute sensitivity to sound. We begin with a very weak sound which the subject fails to hear on repeated trials. We then increase the intensity of the sound. At this second level of stimulation the subject may sometimes hear the sound and sometimes fail to hear it. When we increase the intensity even further, the subject may hear the sound more frequently than before but still miss it part of the time. Finally, we may increase the intensity of stimulation to a level at which the subject never fails to report the presence of the sound. Clearly, then, there is no *one* stimulus value which represents the minimum necessary for a response. For purposes of measurement, it is generally agreed to consider as the *absolute threshold* that stimulus value which yields a response 50 percent of the time, i.e., on half the test trials. It is essential to understand that the absolute threshold is not a *fixed* point on the stimulus scale but rather is inherently variable in time. A single value representing the absolute threshold must necessarily be a *statistical concept*.

Similar considerations apply to estimates of a subject's differential sensitivity. The *differential threshold* is defined as that stimulus difference which gives rise to a judgment of *different* 50 percent of the time. For example, if we present a subject with two tones differing only very little in intensity, he will fail to report a difference most of the time. As we increase the intensity difference between the two sounds so as to obtain a judgment of *different* on half the trials, this difference defines the differential threshold.

There are many variations in the experimental and statistical procedures for the determination of the absolute and differential thresholds, but they all have the same general purpose: to make as good as possible an estimate of that stimulus value which will yield a given judgment—presence vs. absence, same vs. different—on half the trials of a series.

Point of Subjective Equality. One fundamental category of relational judgment is *sameness* vs. *difference*. Sometimes stimuli whose physical characteristics are identical may give rise to a judgment of *different*, and stimuli which differ physically may be judged *same*. Thus, there is no necessary correspondence between physical equality of stimuli and judgments of sameness, nor is there a necessary correspondence between physical differences and judgments of *different*.

For this reason, experiments on discrimination often include an estimate of the *point of subjective equality*. Suppose we present a subject with pairs of stimuli, one member of the pair being fixed and the other member varying from trial to trial, sometimes being equal to the first stimulus, sometimes larger, sometimes smaller. The subject is required to make a judgment of *same* or *different* in response to each pair. In such an experiment the point of subjective equality is defined by that comparison stimulus which is most likely to result in a judgment of *same*. Under many experimental conditions the stimuli most likely to be judged *same* are physically equal ones. Sometimes, however, two stimuli which differ by a certain amount are more likely to be judged *same* than physically equal ones.

Variable and Constant Errors. It is characteristic of psychophysical judgments that they are variable in time. When a subject responds repeatedly to the same stimulus or stimulus configuration,

his judgments may not be the same every time the stimulus is presented. Such variability is especially characteristic of judgments which are rather difficult for the subject; for example, when he is tested with stimuli close to his absolute or differential threshold. There are several probable sources of this variability. The subject's sensitivity may vary from moment to moment, and at any instant the condition of the organism may be more or less favorable to the performance of a fine discrimination. There may, moreover, be slight unavoidable changes in the physical characteristics of the stimulus, for no matter how carefully controlled the instrumentation, a certain margin of fluctuation cannot be eliminated. In addition, such factors as interest and attitude, difficult as it is to measure them, may exercise an important influence on the subject's performance at any moment in time. As a result of these and similar factors, we usually obtain, upon frequent presentations of a given stimulus, a distribution of judgments showing a certain amount of scatter around a most frequent or typical judgment. The degree to which judgments differ from trial to trial provides an index of the amount of *variable error*. The use of the term *error* should not be taken to imply that one judgment is right, and the others wrong. *Error* here simply refers to the extent of fluctuation in judgments.

When two physically equal stimuli, *A* and *B*, are presented to a subject, he does not always perceive them as equal (equally large or intense). He may report a difference even though the physical characteristics of the stimuli are the same. If such judgments are merely a manifestation of variable error, *A* should sometimes be overestimated and sometimes underestimated. Frequently, however, we find a *systematic tendency* on the part of the subject toward overestimation *or* underestimation. Such a systematic tendency is known as *constant error*.

One frequently encountered constant error is the *time error*. If two identical stimuli, *A* and *B*, are presented in succession, we find, for different time intervals between *A* and *B*, systematic tendencies to underestimate or overestimate the second stimulus, *B*. In the case of judgments of loudness, for example, if two equally intense tones are separated from each other by an interval of ½ second, the second tone will usually be judged as less loud than the first. If the two tones are separated by a longer time interval, say, 6 seconds, the

second tone will usually be judged as louder than the first. This tendency to underestimate or overestimate the second of two successive stimuli constitutes the time error. The degree of time error can be plotted as a function of length of time interval between the two tones. The time error has been found to occur with a variety of other stimulus materials, such as judgments of weights and of extent.

Another example of constant error is the *space error*. Judgments may be influenced systematically by the spatial position of the stimuli, whether they are, for instance, on the right or left of the subject.

Again, we must be careful not to be misled by the word *error* in the term *constant error*. Such systematic tendencies as we have described should not be dismissed merely as inaccuracies of judgment which need to be somehow eliminated or corrected. On the contrary, they are of interest to the psychophysicist because they represent a reliable correlation between certain conditions of stimulation and judgment.

Experimental and Quantitative Methods in Psychophysics

It would be impossible to give here an exhaustive survey of psychophysical methods. There are as many psychophysical methods as there are experimental and quantitative procedures for the establishment of lawful relationships between the physical characteristics of the stimulus and the attributes of experience. Nevertheless, there are certain models of experimental procedure and quantification which have had wide application throughout the history of psychophysics to a wide variety of problems and stimulus materials. Frequently these models have been designated as *the* psychophysical methods. In studying these models, we must guard against the misleading impression that there are a few rigidly specified methods for the conduct of psychophysical experiments. The methods to be discussed are *models* in the full sense of the word, illustrating the type of experimental and statistical methods which have been successfully applied to psychophysical problems. Whether or not one works in the framework of the traditional

psychophysical models, each specific experiment must be designed in terms of the questions which will be asked of the data.

Methods for Measuring the Absolute Threshold

We shall first discuss experimental models for the measurement of the absolute threshold. It will be remembered that the absolute threshold is defined as that stimulus value which is detected by a subject on 50 percent of the trials.

The Method of Minimal Changes

A method well adapted for the determination of the absolute threshold is the *method of minimal changes* (known also as the *method of limits*). As the name implies, this method utilizes a series of stimuli, successively differing by small amounts, in order to make an estimate of the absolute threshold.

Basic Procedure. The basic procedure consists of the presentation of a series of stimuli, each differing from the preceding one by a small amount, until a critical change in the subject's judgments occurs. In the determination of the absolute threshold, it is customary to employ two kinds of series: (1) an *ascending series,* in which we start with a stimulus well below the threshold, increasing its magnitude on successive trials until the subject detects the presence of the stimulus; and (2) a *descending series,* in which we start with a stimulus well above the threshold, decreasing its magnitude on successive trials until the subject fails to detect the presence of the stimulus.

For an illustration of this, let us return to the problem of measuring a subject's absolute sensitivity to sound. For the ascending series, we begin with a tone so weak that the subject would fail to report its presence on virtually all the trials. On each successive trial we increase the intensity by a small physical amount until the subject finally reports the presence of the sound. For the descending series, we begin with a tone so loud that the subject reports its presence whenever the tone is sounded. On each successive trial we decrease the intensity by a given physical amount until the subject reports that he can no longer hear a sound.

For the determination of a reliable threshold, several series, both ascending and descending, should be used. In order to prevent the

subject from getting accustomed to a particular serial order, ascending and descending series should be alternated or presented in a prearranged sequence. It is also advisable to vary the *length* of the series. Sometimes we start with a stimulus value well below or above the threshold so that a large number of successive increments or decrements is needed to reach the threshold; sometimes we start much closer to the threshold so that the critical change in judgment occurs after only a few trials. Again, different lengths of series should be alternated according to some predetermined pattern.

Computation of Threshold Value. Each of the series, whether ascending or descending, long or short, yields a threshold value. For

Ascending Series			Descending Series		
Stimulus Value in Physical Units	Judgment		Stimulus Value in Physical Units	Judgment	
10	Absent		25	Present	
11	Absent		24	Present	
12	Absent		23	Present	
13	Absent		22	Present	
14	Absent		21	Present	
15	Absent		20	Present	
16	Present	Change in judgment	19	Present	
			18	Present	
			17	Present	
			16	Absent	Change in judgment

an ascending series, the threshold is given by the mid-point between the last judgment of stimulus *absent* and the first judgment of *present*. For a descending series, the threshold is given by the mid-point between the last judgment of *present* and the first judgment of *absent*. Our records for an ascending and descending series might look somewhat as shown in the above table.

For the ascending series, the absolute threshold is estimated at 15.5, our best estimate of the value at which 50 percent of judgments of *absent* and 50 percent of judgments of *present* would be obtained. Similarly, for the descending series, our best estimate of the absolute threshold would be 16.5. If we had used ten ascending and ten descending series, we would, in the same manner, obtain twenty

estimates of the absolute threshold. Some of these estimates would probably be identical, but in general we must expect a certain amount of variability. Our best single estimate of the absolute threshold is the average of all the threshold values obtained in the individual series. In order to compare the results obtained with ascending and descending series, we compute separate averages for these two types of procedure. If there is a significant difference between the averages obtained for ascending and descending series, we conclude that there is a systematic difference between the subject's responses in the two situations. Similarly, we can compare the results obtained with long and short series, or the judgments during the first half and the second half of the experiment, to gauge the effects of practice and fatigue. In short, the absolute threshold is a statistical average, a dependent variable which can be measured as a function of various experimental conditions.

The Method of Constant Stimuli

Basic Procedure. The method derives its name from the fact that a number of fixed (constant) stimuli is presented to the subject a large number of times, usually in random order. Each time a stimulus is presented the subject makes a judgment. Thus, for each of the stimuli, the percentage of different kinds of judgments may be computed, and an estimate of the threshold is made on the basis of the distributions for all the stimuli.

Suppose we wish to determine the absolute sensitivity to pressure, i.e., the minimum weight that has to be applied to the sensitive receptors of the skin in order to obtain a report of pressure. By means of a delicate instrument (see Chapter 3), graduated pressures can be applied to the skin. To measure the absolute threshold, we select, after preliminary exploration, a series of stimulus values which is likely to give us a distribution of judgments ranging from approximately 5 percent to approximately 95 percent reports of presence of pressure. If we were determining the absolute sensitivity on the tip of the subject's finger, such a series might include pressures of 1, 2, 3, 4, and 5 grams per square millimeter of skin.

Each of these pressures is applied to the skin a large number of times, say, 100. The different stimuli are presented in a random order so that the subject cannot build up an expectation of any particular

sequence. Each time a stimulus is presented, the subject makes a judgment of *present* or *absent* with respect to experienced pressure. Thus, we obtain for each stimulus a percentage of *present* judgments and a percentage of *absent* judgments, the two, of course, always adding up to 100 percent. A possible distribution may be somewhat as shown in the figures on this page.

Remembering our definition of absolute threshold as that stimulus value which yields 50 percent of judgments of *present*, we conclude that the threshold should be estimated as lying between 3 grams (which yields less than 50 percent of *present* judgments)

Stimulus (in grams per sq. mm.)	% Present	% Absent
1	5	95
2	22	78
3	48	52
4	76	24
5	95	5

and 4 grams (which yields more than 50 percent). If we are willing to assume that the increase in judgments of *present* between 3 and 4 grams follows a linear course, we should estimate the absolute threshold at 3.07 grams per square millimeter, i.e., 2/.28 of the distance between 3 and 4 grams. This method of estimating the threshold from the distribution of judgments is that of *linear interpolation* and consists essentially in finding the median (50-percent point) of the distribution of *present* judgments, that stimulus value at which judgments of *present* and *absent* are equally likely.[1]

[1] There are mathematically more accurate methods for estimating the absolute threshold from the judgment distribution. One widely used method, for example, proceeds on the assumption that the distribution of judgments may be best described by a cumulative normal probability curve and weights the percentages accordingly. The theory underlying this particular treatment is that at any moment in time the subject's ability to detect the stimulus depends on a large number of more or less independent factors, some of which are favorable and some unfavorable to the discrimination. These favorable and unfavorable factors combine with each other in accordance with the laws of probability. In the presence of a very weak stimulus, many of the favorable factors would have to occur together to make a correct judgment possible. This happens only very rarely—hence the small percentage of discriminations in the presence of a weak stimulus. With a strong stimulus, only a minimum of the favorable factors

Whether the simple assumption of linearity is made or another method of quantification is used, the absolute threshold is always an estimate of the value which gives an equal proportion of judgments of *present* or *absent.*

Methods for Measuring the Differential Threshold

We now turn to experimental models for the measurement of the differential threshold—the determination of that stimulus difference which is detected by a subject on 50 percent of the presentations.

The Method of Minimal Changes

Basic Procedure. This method, which we described in some detail in connection with the determination of the absolute threshold, may be used for measuring the differential threshold as well. We first decide on a *standard stimulus,* say, a tone or light of fixed intensity. The standard and the comparison stimulus may be presented in succession (e.g., in the case of a pair of tones) or simultaneously (e.g., in brightness discrimination where two fields may be exposed side by side).

The magnitude of the comparison stimulus is changed by a small amount from trial to trial. Again, ascending and descending series are used. For an *ascending series,* we set the comparison stimulus at a value considerably smaller than the standard and then keep increasing it by small regular steps until the subject's judgment of the comparison stimulus changes from *smaller,* to *equal,* and finally to *larger* than the standard. For a *descending series,* we set the comparison stimulus at a value considerably larger than the standard and keep decreasing it by regular steps until the subject's judgments change from *larger,* to *equal,* and finally to *smaller.* As before, the ascending and descending series and series of different lengths are alternated in an orderly fashion.

Suppose our standard, e.g., a sound, has a value of 20 physical

---

need to be present for a correct judgment. This is a very frequent state of affairs—hence the high percentage of correct discriminations in the presence of a strong stimulus.

For a fuller statement of this argument and for a detailed description of the quantitative methods used in the estimation of thresholds, the reader is referred to J. P. Guilford, *Psychometric methods,* New York: McGraw-Hill, 1936, Chap. 6.

units and we vary the comparison stimulus in steps of two units at a time. Typical records obtained with an ascending and a descending series may look somewhat as shown in the table on this page.

Since the judgments are continued beyond the report of equality, we can obtain from each series two threshold values: (1) the *upper threshold* (L_u),[2] i.e., that stimulus which is just noticeably larger than the standard; and (2) the *lower threshold* (L_l), i.e., that

Ascending Series			Descending Series		
Value of Comparison Stimulus in Physical Units	Judgment		Value of Comparison Stimulus in Physical Units	Judgment	
8	Smaller		32	Larger	
10	Smaller		30	Larger	
12	Smaller		28	Larger	
14	Smaller		26	Larger	
16	Smaller		24	Larger	
		Change in judgment	22	Larger	
18	Equal		20	Larger	
20	Equal				Change in judgment
		Change in judgment	18	Equal	
22	Larger		16	Equal	
					Change in judgment
			14	Smaller	

stimulus which is just noticeably smaller than the standard. As in the case of the absolute threshold, we define these thresholds as lying midway between the last judgment of *larger* (or *smaller*) and the first judgment of *equal*. In our ascending series, the upper threshold is found between 20 (last judgment of *equal*) and 22 (first judgment of *larger*), i.e., it is estimated to be 21. In the same series, the lower threshold lies between 16 (last judgment of *smaller*) and 18 (first judgment of *equal*), and is estimated to be 17. Similarly, in the descending series, the upper threshold is located at 19, and the lower threshold at 15.

We are now ready for the determination of the *differential threshold* (DL) itself, i.e., the stimulus difference which is likely to be detected on 50 percent of the trials. Three estimates of the differen-

[2] *L* stands for *limen*, the Latin word for threshold.

tial threshold are possible: (1) the *upper differential threshold* (DL_u)—the difference between the upper threshold and the standard, or $L_u - S$; (2) the *lower differential threshold* (DL_l)—the difference between the standard and the lower threshold or $S - L_l$; and finally (3) the *mean differential threshold* which represents the average of the upper and lower differential thresholds, i.e., $(DL_u + DL_l)/2$. This last value is, of course, the best possible general estimate that can be made of the subject's differential threshold.

Finally, we can now estimate the *point of subjective equality* (*PSE*) as lying midway between the upper threshold and the lower threshold: $\text{PSE} = \frac{L_u + L_l}{2}$. Halfway between the just noticeably smaller stimulus and the just noticeably larger stimulus lies the value most likely to yield a judgment of *equal.*

The Method of Constant Stimulus Differences

Basic Procedure. An alternative procedure for the determination of the differential threshold is provided by the *method of constant stimulus differences.* This procedure is especially well suited for experiments in which (1) it is not convenient to vary the comparison stimulus continuously in small steps, and (2) it is desirable to base the threshold value on a large number of observations.

As the name implies, this method is closely related to the method of constant stimuli which we have already studied. In the method of constant stimulus differences, however, the subject judges *pairs* of stimuli rather than one stimulus at a time. Each stimulus pair consists of the standard stimulus and one comparison stimulus. The standard stimulus is, of course, the same on all trials. A number of different comparison stimuli are used. On each trial, the standard is paired with one of the comparison stimuli. Each time a pair is presented, the subject judges the comparison stimulus as *larger* or *smaller* than the standard. (Sometimes *equal* judgments are permitted.) Each comparison stimulus is paired with the standard the same number of times (e.g., 100 times each). The sequence of pairs is thoroughly scrambled.

The comparison stimuli are chosen after preliminary exploration so as to yield a distribution of judgments ranging from approximately 95 percent *larger* to about 95 percent *smaller.* Stimulus dif-

ferences which are so large as to be always recognized are of little use in determining the subject's differential sensitivity. The final series of comparison stimuli is arranged in even, symmetrical steps around the value of the standard. A comparison stimulus equal to the standard is included. In a weight-lifting experiment with a standard of 200 grams, a typical series of comparison weights might be 188, 192, 196, 200, 204, 208, 212 grams.

On each trial, standard and comparison stimuli are presented (simultaneously or in succession), and the subject judges the comparison stimulus relative to the standard. In some experiments, the subject is allowed to use one of three categories of judgment: *larger*, *smaller*, and *equal*. Frequently, however, the subject is restricted to the use of only two categories and is instructed to guess *larger* or *smaller*, even though he may believe standard and comparison stimuli to be equal. There are several advantages in the use of only two categories: (1) it has been established experimentally that when a subject is forced to guess, he will be more often right than wrong in his guesses; (2) the meaning of the category *equal* is vague and apt to vary considerably from subject to subject; (3) frequency of *equal* judgments may indicate not so much lack of discrimination as the subject's caution and his adoption of very exacting criteria of *larger* and *smaller*. Thus, the exclusion of the *equal* category forces the subject to attune himself as well as he can to small stimulus differences and to exploit, as it were, his capacity for discrimination as much as he can. For these reasons, it is probably well to use the two-category method for the determination of the differential threshold and, in general, in all situations in which the emphasis is on the subject's sensitivity. However, there are situations to which the three-category method is well suited. By means of this method we can investigate the subject's judgment behavior, especially his understanding and use of the *equal* category.

Computation of the Differential Threshold. For each of the *constant* comparison stimuli, we tabulate the percentage of judgments falling into the three (or two) categories of judgment which the subjects had been allowed to use. The specific technique for the computation of the differential threshold varies, of course, with the number of judgment categories.

By way of illustration of the three-category procedure, suppose we are measuring a subject's differential sensitivity to weight. The standard stimulus is a weight of 200 grams, and the comparison stimuli are spaced in 4-gram steps around the standard. A typical distribution of judgments *larger, smaller,* and *equal* may look somewhat as shown in the table below.

A graphic presentation of these results may be found in Fig. 1. The three curves depicting the relative frequency of judgments of *larger, equal,* and *smaller* for different magnitudes of the comparison stimulus are known as *psychometric functions.*

Comparison Stimulus in Grams	% Judged Larger	% Judged Equal	% Judged Smaller
188	5	5	90
192	10	20	70
196	22	35	43
200	38	40	22
204	64	26	10
208	84	10	6
212	92	5	3

From the frequency distributions of judgments, we can estimate the upper threshold, the lower threshold, and the differential threshold. The upper threshold is that value which on 50 percent of the trials yields a judgment of *larger.* Inspection of the table and Fig. 1 shows this value to lie between 200 and 204 grams, more exactly at 201.8 grams.[3] Similarly, we estimate the lower threshold—the value yielding 50 percent of judgments of *smaller*—at 195 grams. Hence, the upper differential threshold is estimated as $201.8 - 200 = 1.8$, and the lower differential threshold as $200 - 195 = 5$, and the mean differential threshold as 3.4. On the average, a difference of 3.4 grams is required for a judgment of difference (*larger* or *smaller*). These results also tell us that in this experiment (as in many others) the subject's differential sensitivity is, as it were, asymmetrical. The average increment required for a judgment of *larger* is less than the average decrement required for a judgment of *smaller.* Finally, the point of subjective equality is most meaningfully defined as that stimulus value which is equally likely to

[3] By linear interpolation between 200 and 204.

give judgments of *larger* and *smaller*. The two psychometric functions cross at 198.5 grams, which is our estimate of the point of subjective equality.

When the subject uses only two judgment categories—*larger* and *smaller*—the treatment of the results is necessarily different. In this case, the two psychometric functions cross at the 50-percent point (for an example, see p. 58). For the two-category procedure,

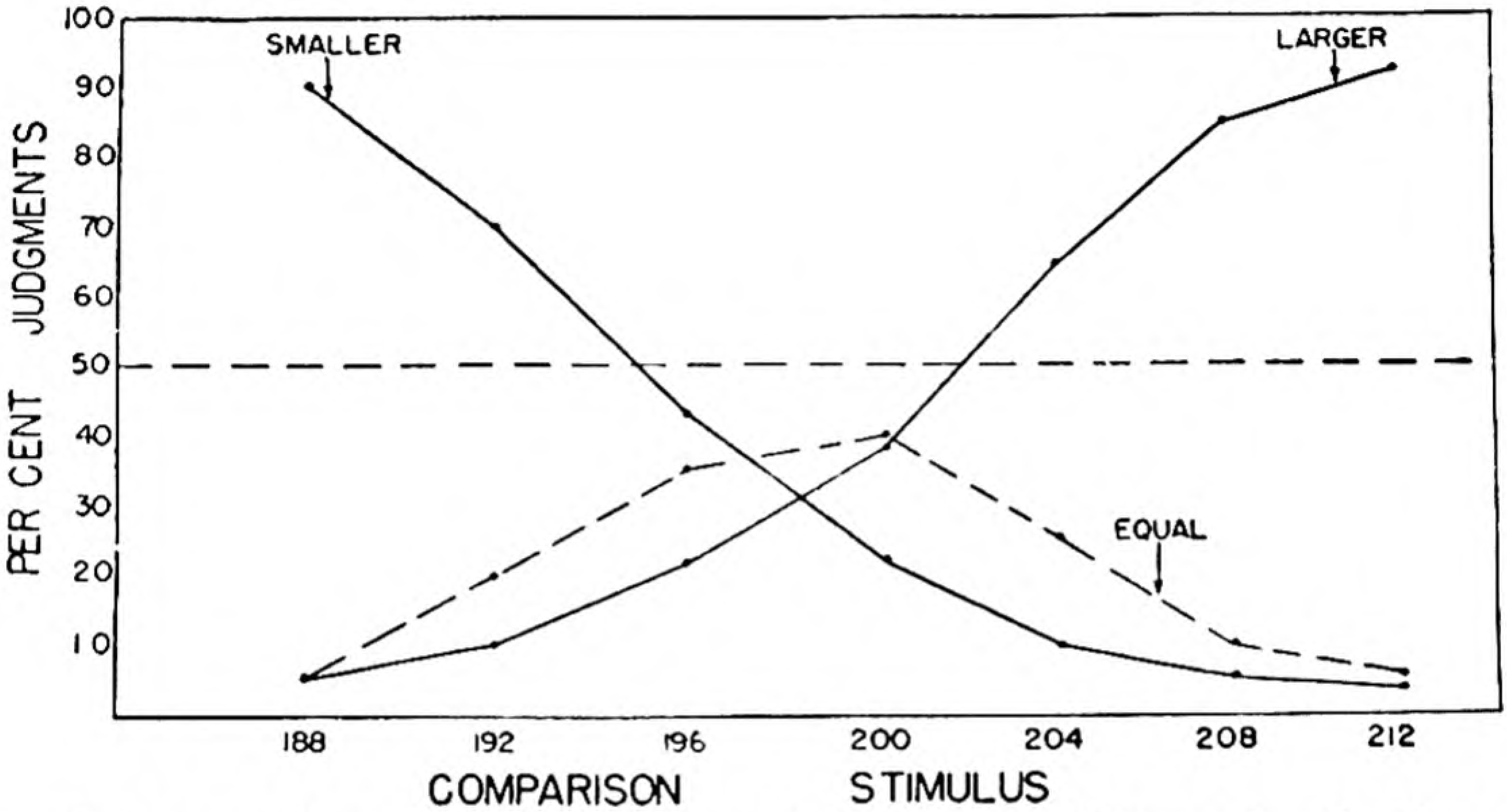


Fig. 1. A Typical Set of Psychometric Functions Obtained with Three Categories of Judgment.

therefore, we *redefine* the differential threshold as that stimulus difference which yields judgments of *different* (*larger* or *smaller*) 75 percent of the time. The point of subjective equality is again defined as that point at which the psychometric functions cross, which in this case is the value yielding 50 percent *larger* judgments and 50 percent *smaller* judgments.

The steepness of the two psychometric functions (their rate of change) depends on the subject's sensitivity. If the subject is very sensitive to each change in the stimulus, the functions will rise (or fall) rapidly; i.e., each change in the comparison stimulus will result in a considerable change in the relative frequency of the two judgments. Conversely, if the subject's sensitivity is poor, any one change in the comparison stimulus will not alter the distribution

of judgments radically, and the psychometric functions rise and fall gradually.

We must emphasize again, as we did in connection with the method of constant stimuli for the determination of the absolute threshold, that the numerical treatment of the data presented here is only one of several possible procedures and, indeed, one of the least precise ones. It is often desirable to fit ideal curves, such as the cumulative normal curve, to the data before computing the threshold values. The procedure used here does, however, illustrate the basic aims and assumptions of the method.

The Method of Average Error

In the determination of absolute and differential thresholds by the methods already discussed, control over changes in the stimulus was entirely in the hands of the experimenter. There are certain problems, however, which may best be attacked by permitting the subject to control the variations in the stimulus. In such cases, the method of average error is used.

Basic Procedure. The most outstanding characteristic of this method is the active role which the subject plays in the procedure. Presented with a fixed standard and a variable comparison stimulus, the subject has the task of adjusting the variable stimulus until it appears equal to the standard. Thus, there may be a line of fixed length and one of adjustable length which the subject manipulates until he is satisfied that the two lines are equal. The subject may adjust a shade of gray until it appears to match a standard sample or change the intensity of a tone until it sounds as loud as a standard. These are just a few examples of a large variety of experimental situations in which the *method of average error* has been employed.

Each adjustment which the subject makes represents a measurement of the point of subjective equality. If a standard line is 200 mm. long and a subject adjusts the variable comparison line once at 198 and once at 202, each of these adjustments provides us with an estimate of the *PSE*. A *reliable* estimate of the *PSE* must be based on a number of such measurements, i.e., the estimate of the *PSE* should be the average of many adjustments made by the subject. Frequently, though by no means always, the average adjustment may be almost equal to the standard. Some of the individual adjustments

may be greater than the standard. Some may be smaller, however, and these positive and negative deviations may cancel each other out; hence, the name of the method—*average error*.

In collecting data by the method of average error, several important precautions must be observed.

1. The direction of adjustment should be varied. On some trials the variable stimulus should be set at a value considerably larger than the standard; on other trials it should be made considerably smaller. In this fashion, the subject has to begin his adjustment toward equality on successive trials from both above and below the standard. The subject typically overshoots his mark and may be allowed to attain equality by successive approximations.
2. Account should be taken of "constant errors." The spatial arrangement of the standard and the stimulus and the temporal sequence of the stimuli (for example, the fact that the comparison stimulus may follow the standard in time) may give rise to systematic judgment tendencies such as the space error and the time error. In arriving at an estimate of the point of subjective equality, such tendencies must be taken into consideration.
3. Care must be exercised to prevent the subject from using—consciously or unconsciously—extraneous cues. For example, if the initial value of the variable comparison stimulus were always the same, the subject might learn a specific movement by which to adjust it to the point of equality. He might learn to turn an intensity control knob through a certain angle. Changing the initial value of the variable from trial to trial will prevent such habits from being formed. Each adjustment must reflect an explicit judgment of equality.

The average of a large number of adjustments represents a good estimate of the point of subjective equality. A significant difference between the value of the fixed standard and the average adjustment (*PSE*) indicates the presence of a systematic "error." If the average adjustment is significantly larger than the standard, the subject overestimates the latter. Conversely, if the average adjustment is significantly smaller than the standard, the latter is underestimated. Significance of such differences must, of course, be determined by statistical analysis. As in other experiments, a difference between standard and *PSE* is significant if it could be expected only very

infrequently on the basis of random fluctuations due to errors of measurement. When a significant difference is established, the source of the systematic error presents a challenge to analysis and further experimental investigation.

The subject's mean adjustment cannot be fully interpreted without considering the scatter of values around the mean. It is important to know not only what the average adjustment to equality is, but also how variable the adjustments are from trial to trial. Some

Adjustments of Variable Tone to Subjective Equality

Trial No.	Ascending	Descending
1	48	51
2	50	47
3	49	50
4	47	48
5	53	49
6	47	52
7	45	46
8	51	48
9	47	52
10	49	47
	Mean Asc.: 48.6	Mean Desc.: 49.0
	Total Mean: 48.8	
	Standard Deviation: 2.1	

index of the variability of the adjustments provides a measure of the *precision* of the subject's judgments. If his equality judgments are precise, they will not vary much from trial to trial. The greater the variability of the adjustments, the less is the *precision* of the subject's judgments.

Method of Computation. The accompanying table summarizes the results of an experiment using the method of average error. The standard stimulus was a tone of fixed intensity (50 physical units). The subject had the task of adjusting the intensity of a variable tone until it appeared equal in loudness to the standard. On half the trials, the comparison tone was set at an intensity considerably above that of the standard, and the subject had to make a downward adjustment to equality. On the other half of the trials, the subject had to adjust the intensity of the comparison stimulus

upward to equality. The mean adjustment is 48.8 which is our estimate of the point of subjective equality. The difference of 1.2 units between the *PSE* and the standard is not statistically significant. A difference of this size would be frequently expected as a result of random fluctuations due to uncontrolled errors of measurement. We cannot conclude that the subject is displaying a systematic tendency toward overestimation or underestimation of the comparison stimulus. Comparison of the mean upward adjustment and mean downward adjustment (48.6 and 49.0, respectively) also fails to reveal a significant difference. Finally, the standard deviation of the distribution of judgments, 2.1, provides us with an index of the precision of the subject's judgments.

The Judgment of Intervals

So far, we have been concerned with situations in which the subject's task has been to judge *presence* or *absence, sameness* or *difference*. We have not asked him to judge the magnitude of the stimuli themselves. Even when the subject adjusts the loudness of a tone so as to match that of a standard, his judgment is focused on the point at which he cannot hear a difference. Yet, a subject finds it possible to judge the magnitude of a stimulus, or the extent of a stimulus difference. He can tell us fairly well when a tone is half as loud or twice as loud as another. As another example, it is fairly easy to adjust a shade of gray until it appears to lie midway between a lighter and a darker gray. Several methods have been specifically designed for the purpose of quantifying these judgments of magnitudes and intervals.

The Method of Fractionation

In this method, the subject has the task of varying the value of one stimulus (the variable stimulus) until it appears to have half the value of another stimulus (the fixed standard). For example, the subject may be provided with a tone of fixed intensity and then be asked to adjust the intensity of the variable until it sounds half as loud as the standard.[4] The relation between standard and variable can, of course, be reversed, and the subject be required to

[4] See pp. 61 f. for an application of this method to the construction of sensory scales.

adjust the variable until it appears to have twice the value of the standard. Frequent experimental tests have established subjects' ability to make such judgments of relation with satisfactory consistency. This procedure may be varied by using a series of fixed comparison stimuli which the subject judges as to whether or not they are twice (one-half) the value of the stimulus. Thus, both the method of average error and the method of constant stimulus differences can be adapted to problems of fractionation.

The Method of Bisection

This procedure is closely related to the method of fractionation The subject's task is to adjust a variable stimulus until it appears to fall midway between two fixed stimulus values. Thus, given stimuli S_1 and S_3, say, tones of fixed intensity, the subject has to adjust a variable stimulus, S_2, say a tone of variable intensity, until it appears to him that the distance between S_1 and S_2 equals the distance between S_2 and S_3. From a large series of such adjustments, an estimate of the subjective mid-point of the interval S_1 to S_3 can be made.

The Method of Equal-Appearing Intervals

If a subject is required to grade a series of stimuli so that they appear equally distant from each other, we speak of the method of *equal-appearing intervals*. The method has probably found its best known application in the construction of the Thurstone attitude scale. The judge is presented with a large number of statements concerning an institution, national group, or other issue, and sorts them into a series of categories which he judges to be equally spaced in favorableness. On the basis of such judgments, scale values may be computed for each of the statements.

We mention the methods of fractionation, bisection, and equal-appearing intervals only briefly since they are used less frequently than the other methods we have discussed and have found their main application in advanced psychophysical research. They serve well to illustrate, however, the complexity of relational judgments which the human subject can make with good reliability. The analysis of sensory discrimination is not confined to the measurement of thresholds.

Comparison of Experimental Procedures

We often hear about *the* psychophysical methods—the method of minimal changes, the method of constant stimuli (and constant stimulus differences), and the method of average error. We have emphasized that these methods should be regarded as generalized models rather than as cut-and-dried procedures which must be slavishly followed by the experimenter. They do, however, embody the most common experimental approaches to the measurement of discrimination—absolute and differential. Let us now briefly examine their main similarities and differences.

Subject's Task. In a psychophysical experiment, the subject may either be presented with stimuli and be confined to making a judgment or he may actively manipulate a stimulus until he has reached the criterion prescribed by the experimenter. In the method of minimal changes and the constant methods, the subject's role is usually passive but he actively manipulates a stimulus in the method of average error. Active manipulation of the materials may help to maintain the subject's interest and motivation. On the other hand, it may complicate the evaluation of the data, for the subject's skill in handling the apparatus, e.g., the steadiness of his hand in making an adjustment to equality, may introduce an uncertain amount of variability into the results. In some experiments, on the other hand, we may be specifically interested in such motor errors and desire to measure them.

Fixed vs. Variable Stimuli. In the constant methods, the stimuli are fixed and their values cannot be changed during an experimental series. In the method of average error, the comparison stimulus is continuously variable at the subject's discretion. The method of minimal changes occupies a position intermediate between these poles. The stimulus is varied by fixed amounts but the *extent* of variation may vary from series to series. Continuous variation lends greater flexibility to the experimental procedure. On the other hand, fixed stimuli make for greater reliability, for a larger number of observations is obtained for any one stimulus value.

Number of Observations for Determination of Critical Stimulus Values. In the method of minimal changes, each series yields a threshold value; in the method of average error, each adjustment

yields a point of subjective equality. Good estimates of thresholds and *PSE*'s are made by averaging a number of such observations. In the constant methods, on the other hand, any one judgment of *different* does not yield a measure of the magnitude of the threshold. Only *one* threshold value and *one PSE* are obtained, but these are based on a very large number of judgments and are inferred from the total distribution of judgments.

Randomness of Stimulus Sequence. In the constant methods, stimuli follow each other in a random order which the subject is unable to predict. In the method of minimal changes, by its very nature, stimuli follow each other in a systematically increasing or decreasing series. Even though the experimenter is careful to use ascending and descending series, the subject inevitably builds up expectations about the nature of the series and is in a state of expectancy for a change in judgment (from *absent* to *present*, from *larger* to *equal*, from *smaller* to *equal*, etc.). In his state of readiness, the subject may easily "jump the gun," report a change earlier than he would have in the absence of such expectation. Even with the most coöperative subjects, such errors can hardly be avoided. As far as freedom from the factor of suggestion is concerned, the constant methods are clearly superior.

Among the various psychophysical models, the experimenter must take his choice according to the emphases demanded by his particular problem. Above all, he must feel free to combine, adjust, and modify these procedures so as to get the best possible answer to the questions which he wishes to ask of his experimental results.

General References

Boring, E. G., The control of attitude in psychophysical experiments, *Psychol. Rev.*, 1920, *27*:440-452.

Brown, W., and Thomson, G. H., *The essentials of mental measurement*, Cambridge: University Press, 1921.

George, S. S., Attitude in relation to the psychophysical judgment, *Amer J Psychol.*, 1917, *28*:1-37.

Guilford, J. P., *Psychometric methods*, New York: McGraw-Hill Book Co., 1936, Part I.

Irwin, F., Psychophysical measurement methods, *Psychol. Bull.*, 1935, 32.140-171.

Johannsen, D. E., *The principles of psychophysics, with laboratory exercises,* Saratoga Springs, New York: the author, 1941.

Kellogg, W. N., An experimental comparison of psychophysical methods, *Archiv. Psychol.*, 1929, No. 106.

Thurstone, L. L., Psychophysical methods. In *Methods of psychology,* T. G. Andrews (ed.), New York: John Wiley and Sons, 1948.

Thurstone, L. L., and Chave, E. J., *The measurement of attitude: a psychophysical method and some experiments with a scale for measuring attitude toward the church,* Chicago: University of Chicago Press, 1929.

Titchener, E. B., *Experimental psychology,* New York: The Macmillan Co., 1905, Vol. II, Parts I and II.

Urban, F. M., The application of statistical methods to problems of psychophysics, *Exper. Stud. in Psychol. and Pedag.*, University of Pennsylvania, 1908, No. 3.

Woodworth, R. S., *Experimental Psychology,* New York: Henry Holt and Co., 1938, Chaps. 17 and 18.

»»» 3 «««

CUTANEOUS SENSITIVITY

IN THIS chapter we begin the study of the psychophysics of sensory relations. We have defined psychophysics as the investigation of lawful relationships between the measurable characteristics of the stimulus, on the one hand, and response characteristics, on the other. The general field of psychophysics naturally divides itself into areas according to the separate sense modalities—vision, hearing, smell, taste, and the cutaneous senses. Although the concept of a sense modality may be intuitively obvious, it merits a brief formal consideration.

The Concept of a Sense Modality. The concept of a sense modality begins with the fact that the organism is equipped with specialized receptor systems. These receptor systems are differentially sensitive to various forms of energy changes in the environment of the organism. Thus, the eye is especially sensitive to radiant energy, and the ear is maximally sensitive to sound waves. It must not be thought that these sensory systems are insensitive to other forms of energy change. The receptors in the eye may be activated by pressure, and the auditory apparatus by electrical stimulation. The sensory systems, however, are specialized in that they are constructed to provide information about the environment through their response to a particular form of energy. This specialization is, moreover, continued anatomically by the spatial arrangement of the central representation of these receptor systems. Thus, the optic nerve fibers have their central representation in a common region of the brain called the visual cortex. Similarly, the cutaneous senses have a delimited area of central representation.

The concept of a sense modality involves more than this. We must consider, for a moment, the characteristics of sensory experience. Varied as these characteristics are, they form separate groups

of continua. All the hues and brightnesses of vision form one group. Noises, musical tones, and speech sounds all form a different group. *Within* each of these groups there are continuous transitions; from group to group, however, there are marked discontinuities. It is in terms of such discontinuities that the sense modalities are marked off from each other. It is the study of the characteristics of experience, above all, which enables us to distinguish among the various skin senses.

The Cutaneous Senses. Although the skin is richly supplied with nerve endings and receptor cells, there is no special grouping of them into a single sense organ as we find with the eye and the ear. The skin senses do not constitute sensory systems which can be independently defined both in terms of specific physical stimuli to which they respond *and* in terms of distinct anatomical and functional systems. Rather, the skin senses must be defined and distinguished from each other in terms of varieties or *dimensions* of response to stimulation of the skin. Once the types or dimensions of response have been defined, a search for neurological correlates and specific receptors can be carried on.

On the basis of response analysis, four skin senses have been distinguished: the pressure sense, the pain sense, the warmth sense, and the cold sense. Note again that the breakdown is not initially based on the anatomically separate sensory systems but on categories of response. The search for specialized types of receptors is still in progress.

Punctate Distribution of Sensitivity. Let us stimulate the skin with a small camel's hair, a sharp needle, and the end of a rod whose temperature we control. As we successively apply these stimuli to different parts of the skin, we do not find uniform sensitivity to them. Rather, some areas or spots are sensitive to slight contact but not to changes in temperature. Again, other areas or spots are sensitive to temperature changes but not to pressure. Even areas or spots sensitive to temperature break down into "warm spots" and "cold spots." We speak, therefore, of punctate distribution of skin sensitivity. *Punctate sensitivity* implies not only markedly different thresholds in different areas but also the existence of distinct spots sensitive to different kinds of stimuli.

We speak of *mapping* the skin when we determine the distribu-

tion of punctate sensitivity to a given kind of stimulus. Usually, a grid of indelible ink is stamped on the skin of the subject to help locate the sensitive spots. These grids provide fairly good means of localizing the stimulus when we try to map the same skin areas repeatedly in order to determine the stability of the sensitive spots. Such a grid does not provide us, however, with perfectly stable coördinates for the mapping of the skin, since the skin is pliable and the location of the grid may shift. Other problems of mapping will be treated in connection with the individual cutaneous senses.

The Pressure Sense

Punctate Distribution. In mapping the spots sensitive to pressure, we usually employ as a stimulus contact of a camel's or horse's hair with the skin. Application of the end of a hair exerts a force on the skin and deforms it. Deformation of the skin (which may also be brought about by pulling rather than depressing the skin) is the effective stimulus for pressure. The force applied may be varied by using hairs of different stiffness and/or length. These hairs may be calibrated by determining the weight in one pan of a balance which the hair will just lift when it is applied to the other pan. By measuring the diameter of the hair, the stimulus may then be expressed as grams per millimeter or grams per square millimeter.

There are two questions which we attempt to answer in mapping the pressure spots: first, for a stimulus of given size and stiffness, how many spots can be found in a given region of the skin? Second, what is the absolute threshold for a given region? This latter measurement is made by employing a graduated series of hairs. The number of pressure spots in any given region will depend on the specific force applied.

Using a stimulus of constant force, pressure spot distributions, such as those shown in Fig. 2, are obtained. It will be noted that pressure spots are found predominantly, but not exclusively, in the hairy regions of the body.

Different parts of the body vary both in the number of spots per unit area and in absolute threshold. For example, the fingertips and the lips are highly sensitive to pressure, with over 100 spots per square centimeter. In some areas, however, such as the cornea of the eye, there are no pressure spots. In general, the number of

pressure spots is greatest toward the extremities of the body. Moreover, absolute sensitivity tends to increase with the density of the spots. There are, then, some regions, especially toward the extremities, which are characterized by both great density of pressure spots and low absolute thresholds.

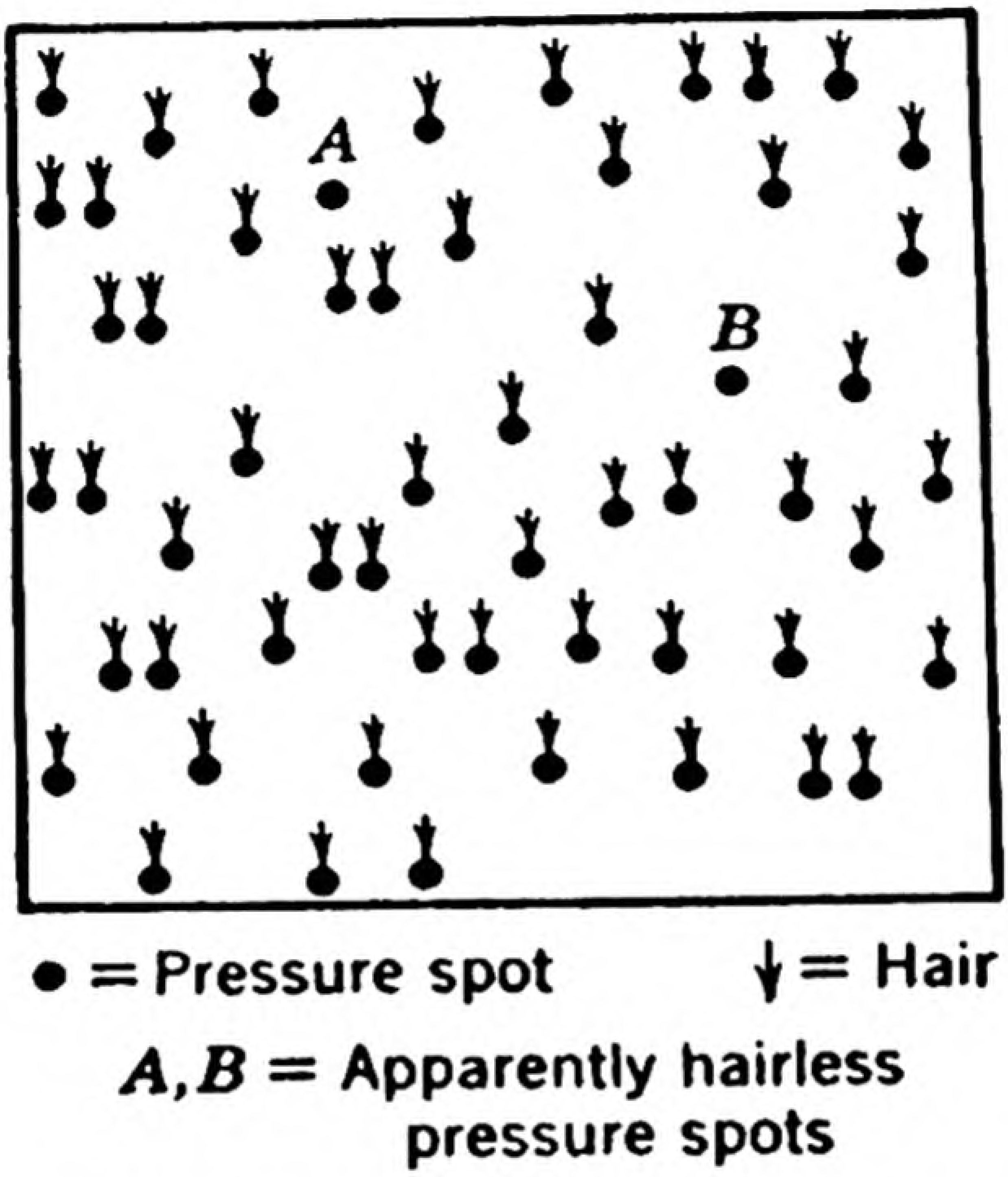


Fig. 2. A Map of Pressure Spots on the Middle Forearm. Note the close but not perfect correspondence between pressure spots and the presence of hairs. (Reproduced by permission from *Foundations of Psychology* by Boring, Langfeld and Weld, p. 362, published by John Wiley & Sons, Inc., 1948. After Strughold, 1925.)

The value of the absolute threshold may be lower than 1 gram per square millimeter (tip of finger) and as large or larger than 100 grams per square millimeter on the sole of the foot. The threshold for pressure depends, of course, upon the rate of application of the force. If the force is rapidly applied, the threshold value is small; if it is slowly applied, it is larger. One further fact concern-

ing pressure sensitivity may be noted. If the area of contact with the stimulus is very small (hair stimuli), the absolute threshold is independent of the extent of the contact area, provided we express the thresholds in terms of grams per millimeter (force per linear size of stimulus). Thus, after obtaining the absolute threshold with

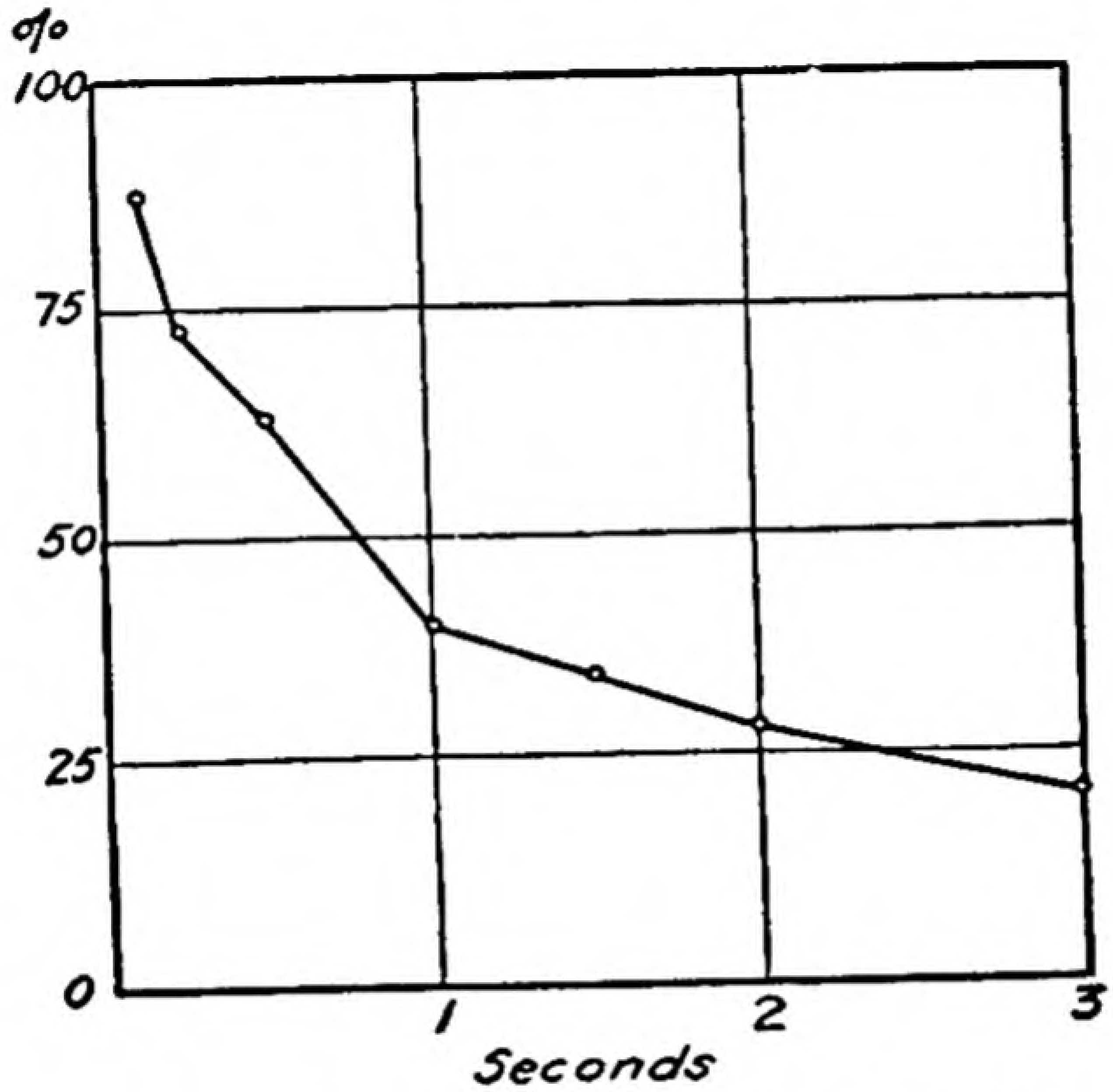


FIG. 3. A Curve Showing the Course of Adaptation to Pressure. For a full explanation of the experimental procedure used to obtain this curve, see text. (From R. S. Woodworth, *Experimental psychology,* 1938, p. 468, by permission of Henry Holt and Company, Inc. After Von Frey and Goldmann, 1915.)

a given hair, if we now use a hair with double the diameter, we need to double the force in order to elicit a response.

Adaptation. The phenomenon of sensory adaptation refers to the fact that with the steady application of a stimulus there is a gradual decline in the sensory effect. Sense organs are most sensitive to rapid changes in the environment; in some modalities, a steady stimulus after a while fails to evoke its initial response. There is some evidence that for cutaneous pressure, the sensory effect

coincides with the time during which the skin is being deformed. When the stimulus has deformed the skin and come to rest, the experience of pressure tends to disappear.

The decline in the sensory effect is difficult to measure quantitatively. In one experiment, the decline in the sensory effect was measured by an indirect procedure, using momentary pressures in one area of the skin for comparison with the declining sensory effect of a continuing stimulus in another area. Thus, by adjusting a momentary pressure in intensity to the pressure aroused by the steadily continuing stimulus, it was possible to determine the course of the decline in the sensory effect of the latter. Fig. 3 shows that considerable adaptation has occurred in a very few seconds.

This experiment illustrates an important type of procedure used in the study of sensory phenomena. We determine the effect of one stimulus in terms of the effects of another and different stimulus by requiring the subject to match these stimuli with respect to some single aspect. In this manner, standards of known value can be used in the study of sensory effects. The visibility curves and the equal-loudness contours illustrate, in other sense modalities, the same type of procedure (for discussion of these, see pp. 64 and 115).

The Pain Sense

Punctate Distribution. The free nerve endings in the skin may be injured in a variety of ways: they may be pierced by a needle or some other sharp instrument; they may be burned by the application of excessive heat; or they may be harmed by chemical agents. Such injury of the free nerve endings arouses the experience of pain. Pain, then, may be due to mechanical, thermal, and chemical action on the skin. In addition, as in the case of the other senses, electric shock may be a very effective stimulus to pain.

Pain spots are more numerous than pressure or temperature spots. Although different parts of the body vary considerably in density of pain spots, there are very few skin areas entirely devoid of them. In the vicinity of those nerves and blood vessels which are close to the surface of the skin, pain spots are particularly numerous. This is, of course, fortunate because pain serves as a signal of danger. The position of the pain spots is independent of the position of the pressure spots. One notable instance is the center of the cornea which

gives rise only to pain but not to pressure, warmth, or cold. The independence of pain and pressure spots can obviously be verified only by the use of weak, small (microscopic or needle-point) stimuli which may produce either pain or pressure, depending on the location. If the stimulus is intense or large in area, the pliability of the skin will insure that both pain and pressure spots are stimulated.

Just as the distribution of pain spots varies from region to region, so do the absolute thresholds for pain. With a mechanical stimulus, the threshold for pain in the center of the cornea may be as low as 0.2 gram per square millimeter. On the other hand, on the tip of the finger a force as large as 300 grams per square millimeter is required to elicit pain. Note that the magnitude of the thresholds for pain and pressure do not vary together in a given region of the body. For example, the tip of the finger, relatively insensitive to pain, is highly sensitive to pressure. Again, as in the case of pressure, great density of pain spots is accompanied by low absolute thresholds.

Adaptation. Common experience suggests that pain does not show the phenomenon of sensory adaptation. Nevertheless, if the stimulus, after application to the skin, is carefully kept unchanged, there may result a pain experience which gradually declines in time. Under many conditions the stimulus is not kept unchanged. Injury may reoccur when the skin moves and, thus, with this new stimulus, the pain will reoccur. Failure to keep the stimulus unchanged may account for the lack of pain adaptation in daily experience.

As compared with adaptation to pressure, adaptation to pain is slow. Even under controlled conditions, it may take from about 10 seconds to about 10 minutes for complete adaptation. These times vary with the nature of the stimulus applied and the particular spot stimulated.

The Temperature Senses

The regions sensitive to cold do not coincide with those sensitive to warmth. It would be possible, therefore, to deal separately with a cold sense and a warmth sense. In view of many functional similarities, however, and because both cold and warm receptors are involved in certain complex sensory effects, the two types of sensitivity to temperature will be treated together.

Physiological Zero. If we apply a stimulus of variable tem-

perature to a specific region of the skin, we discover a particular temperature value or range of temperature values at which the stimulus fails to produce a sensory effect. This point of indifference is defined by the temperature of the skin at the point of stimulation. In other words, if a thermal stimulus is equal to, or only very little different from, the temperature of the skin, there is no experience of warmth or cold. This point or region of indifference is known as the *physiological zero.* In order to be effective, a thermal stimulus must represent a *change from physiological zero.* As for the specific physical value of physiological zero, it varies for different parts of the body and is, of course, a function of the temperature in the environment. Under conditions of normal room temperature, physiological zero, for exposed parts of the skin, is in the vicinity of 33° C. Effective stimuli lower than physiological zero give rise to experiences of cold, those higher than physiological zero to experiences of warmth.

The effectiveness of a thermal stimulus depends, however, not only on its difference from physiological zero, but also on some of its other physical characteristics, such as its specific heat and its efficiency as a thermal conductor. Let us take two objects of equal temperature, both somewhat below physiological zero: a piece of metal which is a good conductor and a piece of cotton which is a poor conductor. Both will at first feel cool, because they are below physiological zero. The metal, however, will feel cooler than the cotton. Metal conducts heat away from the skin more rapidly than does cotton.

It is not necessary to go much above or below physiological zero to arouse experiences of warmth or cold. If the physiological zero of the skin is near normal, e.g. 32° C., an increase or decrease of about 0.05° C. will result in an experience of warmth or cold respectively. Thus, there is a *neutral zone* of 0.10° C. around the physiological zero. Physiological zero, then, is not a point on the scale of temperature but a small region.

Adaptation and Shifts in Physiological Zero. Whether or not a stimulus of fixed intensity arouses warmth or cold depends on the position of the physiological zero. Physiological zero, however, is not a fixed value for any given region of the skin; it may go up or down as a result of the continued application of a stimulus. Such

continued application of a stimulus leads, as we remember, to the phenomenon of sensory adaptation. In the case of continued thermal stimulation, two distinct adaptation effects can be established: first of all, the experience of warmth or cold declines to a weak level. Thus, if the hand (the physiological zero of which is 32° C.) is immersed into a bath of about 20° C., the experienced coldness of the water continually declines, though it may not completely disappear. In addition, the position of the neutral zone along the physical temperature scale has shifted radically. If, after adapting the hand to 20° C., we apply various thermal stimuli to it, we not only find that temperatures below 20° C. feel cold but also that temperatures between 20° and 32° C., which felt cold before adaptation, now feel warm. A shift in the physiological zero involves both warmth and cold. The neutral zone does not expand so as to cover the range from 20° to 32° C. The neutral zone does, however, enlarge to about a degree when adaptation is to extreme temperatures.

Although adaptation to thermal stimuli is marked, it is not always complete. Complete adaptation does occur with punctiform stimulation. It also occurs with areal stimulation, e.g., immersion of a finger into water when the adapting temperature does not deviate too much from physiological zero. If areal stimuli with extreme adapting temperatures are employed, adaptation is only partial. If the finger is immersed into water of 10° C., the experience of cold continues for a long time. If the temperature of the water is above 40° C., the experience of warmth or heat may continue indefinitely. Complete adaptation fails to occur for temperatures below about 15° and above about 40° C. with areal stimulation by water. Not only the extent of adaptation but also the time to attain a steady level of adaptation varies with the temperature of the adapting stimulus. In general, the more extreme the adapting temperature, the more time is required for adaptation.

When extreme temperatures are used, new qualities of experience are aroused. At one end of the scale, extreme cold shades into pain; at the other end, warmth changes to heat, and finally, a burning painful experience is reported. Painful cold is reported at about 5° C., heat at about 43° C., burning heat at about 47° C.

Punctate Distribution. As in the case of pressure and pain, we can determine the location of the warm and cold spots over the sur-

face of the body. The end of a blunt metal rod of a given temperature may be conveniently used for stimulation. It is, of course, extremely important that for a given determination the temperature of the rod be constant. This may be achieved by immersing most of the rod in a bath of constant temperature. In mapping the warm spots, a stimulus device which is electrically heated is convenient.

Again, as in the case of pressure and pain, the density of cold and warm spots differs for various parts of the body. The range of the number of cold and warm spots per square centimeter is much less than that for pressure or pain. The absolute number of temperature spots found depends, of course, on the temperature of the stimulus with which the skin is explored. If the thermal stimulus is made intense enough, there is hardly any area of the external skin which will not produce experiences of warmth or cold. This may, however, be due to the fact that a very intense stimulus spreads over a wide area of the skin.

Since the skin conducts heat, and the stimulus, therefore, cannot be confined to an exactly circumscribed region of the skin, it may be that thermal sensitivity depends upon the *concentration* of receptors in a given area. If the concentration is great, every application of a thermal stimulus will reach a sensitive receptor. If there is a low concentration of receptors, a given stimulus has a smaller chance of activating a receptor. Because of the changing concentration of thermal receptors, there arise "hills and valleys," areas of greater and less sensitivity.

The number of cold and warm spots is smaller per unit area than is the case for the other two senses. There are areas of the body that are insensitive to cold or warmth. Stimulation of the center of the cornea of the eye produces only experiences of pain. The outer margin of the cornea, however, is sensitive to both painful and cold stimuli, but not to warm ones. Certain areas of the throat are also insensitive to temperature change.

As we have mentioned before, the absolute threshold to cold or warmth is small, about 0.05° C. change in temperature when the physiological zero is near normal. Under adaptation, however, the threshold value may be increased to as much or more than half a degree. It is interesting to note that the thresholds for cold and

warmth are approximately equal on most regions of the external skin.

So far, our discussion has implied that stimuli above the physiological zero give rise to warmth, and stimuli below the physiological zero give rise to cold. This is generally true, of course, but there are interesting exceptions. Let us suppose that we have already located a cold spot. If we apply a stimulus above the physiological zero to that spot, the subject may report an experience of cold. Since the relation between the temperature of the stimulus and the report typically given is reversed, we call this phenomenon *paradoxical cold.* A stimulus of fairly high temperature (about 45° C.) has to be applied to a cold spot in order to evoke paradoxical cold. The parallel phenomenon of *paradoxical warmth,* resulting from the application of a low-temperature stimulus to a warm spot, is much more difficult to demonstrate.

Paradoxical cold cannot be aroused by an areal (non-punctiform) stimulus because the warmth receptors would also be stimulated in that case. An areal stimulus of about 45° C. must stimulate both warmth and cold receptors. Since heat is experienced at that temperature, the hypothesis has been proposed that heat results from simultaneous stimulation of these two types of receptors. This hypothesis has been verified by the following experiment. We choose a warm spot and apply to it a stimulus of about 38° C., which under ordinary circumstances arouses an experience of mild warmth. Close to this warm spot we select a cold spot and use a stimulus of about 20° C., which by itself is ordinarily experienced as cold. When the two spots are stimulated together, the subject will frequently report an experience of heat. This experiment is called the *synthetic heat experiment,* since heat is produced by two stimuli, neither of which is hot.

Whatever the quality of experience aroused, temperature spots are somewhat unstable. Suppose we select a circumscribed area of the skin with a grid of indelible ink upon it, and then carefully map the cold and warm spots at two temperatures. A day later we remap the same area with stimuli of the same temperatures. We find that while some of the warm and cold spots have disappeared, many spots have remained invariant, while again some formerly insensitive areas are now responsive. Although the distribution of

the receptors is punctate (the receptor cells and free nerve endings are discrete), nevertheless we obtain instability in mapping because the skin itself conducts heat over areas which do and areas which do not contain sensitive cells. The extent to which the effect of a given stimulus will spread is variable, and it may reach different numbers of receptors from one application to another.

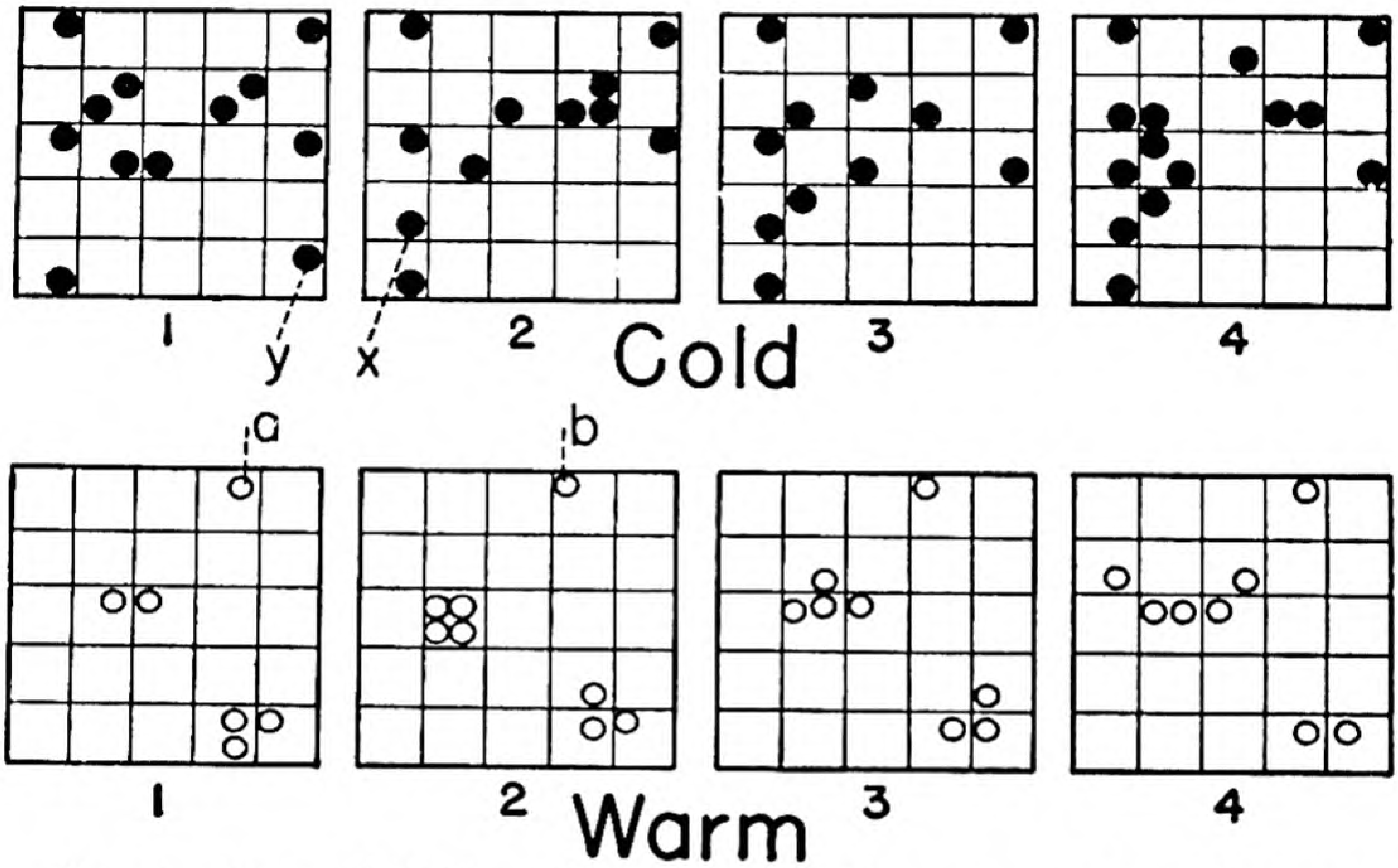


FIG. 4. Maps of Cold and Warm Spots. The same area on the upper arm was plotted four times in the course of one week. Note that the distributions of spots changed somewhat from one mapping to another, but also show considerable consistency. (From K. M. Dallenbach, The temperature spots and end organs, *Amer. J. Psychol.*, 1927, 39:417, by permission of the journal and the author.)

Moreover, slight changes in the temperature of the skin and slight shifts in the physiological zero are likely to lead to a change in the mapped distribution of sensory spots. Even the most careful experimentation cannot avoid the occurrence of such changes from day to day.

As Fig. 4 shows, the changes in the distribution of the spots are not haphazard. There are many stable spots and some that have shifted. Such changes are precisely what we would expect in view of the previous discussion.

Cutaneous Receptors

The distinction among four skin senses has naturally led to a search for specialized receptors. One procedure has been to correlate the distributions of sensory spots with the distributions of receptor-type cells and neural elements in the skin. A second, and more direct procedure, consists of the histological examination of a skin area that has been carefully mapped for sensory spots. The histologist has described and classified many elements which are potential cutaneous receptors, but the correlation between these histological elements and observed sensory effects is equivocal.

In the early work evidence seemed to point toward specialized receptors for the four senses. As experimental work accumulated, however, it was found that, in general, there was no one-to-one correspondence between the four types of cutaneous experience on the one hand, and four types of receptors on the other. For example, the margin of the cornea is sensitive only to painful and cold stimuli. Both free nerve endings and Krause's end bulbs have been found in that area. There the correlation appears to be between pain and free nerve endings and between cold and Krause's end bulbs. When this hypothesis was tested in other areas of the skin, histological examination of cold-sensitive spots failed to show the presence of these end bulbs in some cold-sensitive areas. On the other hand, the correlation between pain and free nerve endings seems established, for the free nerve endings are the only receptor elements frequent enough to account for the great density of pain spots. In addition, direct histological tests have supported this conclusion. Only one other correlation seems certain, and that is between the hair follicles or bulbs and the experience of pressure in the hairy regions of the body. The frequently mentioned correlation between Meissner's corpuscles and pressure in the hairless regions of the body is, on the other hand, doubtful. There has been little, if any, agreement on the end organ for warmth.

The equivocal evidence concerning specialized receptors, the increasing number of end bulbs and corpuscles which have been identified, and the ubiquitous presence of free nerve endings have led some investigators to believe that it is only the latter which are always functional in cutaneous sensitivity. Although the free nerve

endings appear to be structurally alike, they may form different functional systems.

Recently, a more fruitful line of investigation has been the correlation between cutaneous experiences and various types of nerve fibers. Physiologists have classified the nerve fibers of a cutaneous nerve into three types (*A*, *B*, and *C*) in terms of certain functional and anatomical characteristics. From experimental and pathological evidence, and from electrical records of nerve action, a certain degree of correspondence between fiber types and types of cutaneous experiences has become evident. Whatever the specific receptors which respond differentially to various types of stimuli, this differentiation is maintained by the fibers of the cutaneous nerves.

Experiment I

Demonstration of Thermal Adaptation

Purpose. The purpose of this classical experiment (which was first performed more than a hundred years ago) is to demonstrate the shift in the physiological zero as a result of continuous exposure to a thermal stimulus.

Materials. Three beakers or bowls, water, ice, heater (or a source of warm water), and a thermometer are required.

Procedure. We first wish to show that the physiological zero of one skin surface shifts due to exposure to warm water, while the physiological zero of another surface shifts in the opposite direction due to exposure to cold water.

The temperature of the water in one bowl (*B*) is adjusted until it feels neutral when the subject immerses his fingers in it. This will probably occur at about 30° C. The temperature of the water in one of the other bowls (*A*) is lowered to 20° C. Add ice if necessary. The temperature of the third bowl (*C*) is raised to 40° C.

The subject immerses the fingers of his left hand in bowl *A* and the fingers of his right hand in bowl *C*. He keeps them in the bowls for 2 minutes. At the end of that time, he dips the fingers of both hands simultaneously in bowl *B*. Bowl *B* contains the water which previously had felt neutral. The subject now reports how the water feels to each hand.

As a further demonstration of the fact that the physiological zero simultaneously shifted for both thermal senses, the following procedure may be used. Only two bowls are required. The temperature in one of them is adjusted to 25° C., and the temperature of the other to 15° C. The subject momentarily dips his fingers into the first bowl (25° C.) and reports his thermal experience. He then immerses the fingers of the same hand

into the second bowl (15° C.) and keeps them there for 2 minutes. At the end of this period he again tests the water in the first bowl (25° C.).

Experiment II
Mapping Cutaneous Sense Spots

Purpose. To map the four types of sensory spots (pressure, pain, cold, and warm spots) in a circumscribed area of the skin.

Pressure Spots

Materials. An *esthesiometer,* i.e., a horse's or camel's hair attached to a handle, is used for the application of the pressure stimulus. An area of the skin is delineated for plotting by means of a grid stamp. A safety razor is used for shaving off hair previous to mapping the pressure spots and a very light grade of sandpaper for rubbing off the outer layer of skin. The map of pressure spots is recorded on graph paper.

Procedure. Two skin areas are selected: one from the back of the hand, and the other from the dorsal part of the upper arm. For each of the two areas, the following procedure is used. The area is first freed of hair, washed with warm water and lightly rubbed with the sandpaper in order to remove the outer, dead layer of skin. The grid is carefully stamped on the skin with an attempt to make the lines as clear and sharp as possible. A piece of graph paper is marked off with a grid corresponding to that stamped on the skin. Before beginning the actual mapping of the area, the experimenter should practice using the esthesiometer so that he is able to bend the hair to the same degree on each trial. In this manner the intensity of the stimulus will be kept constant. Only under this condition can the maps be meaningfully interpreted. During this practice period, the subject should attempt to establish a consistent criterion for reporting the presence of pressure.

The area is then mapped by applying the esthesiometer to each square of the grid. After each application of the esthesiometer, the subject reports whether or not he felt the pressure. Whenever pressure is reported, the experimenter marks on the graph paper the square corresponding to the one from which the subject's response was obtained. This is continued until all the squares of the grid have been stimulated.

Pain Spots

Materials. Since the same areas of the skin are to be mapped, few new materials are required. The stimulus is applied by means of an *algometer* which is simply a sharp, slender needle attached to a handle. A steel phonograph needle mounted in a handle serves as well. A wet

cloth or sponge is needed to keep the skin moist during the experiment.

Procedure. Before mapping the pain spots, the experimenter should practice the use of the algometer. It will be discovered that very little pressure is needed to elicit pain. The experimenter must be able to apply the needle with the same intensity on all trials. After the mapping has begun, special care must be taken not to increase the intensity of the stimulus if a few successive trials fail to produce a response. Again, each square of the grid is tested, and a record made on the map. The skin should be kept moist throughout the mapping.

Warm Spots

Materials. The warm stimuli are applied by means of temperature cylinders or brass rods. A beaker with water and a heater are required to keep the stimulaters at a constant temperature. A centigrade thermometer is used to measure the temperature.

Procedure. We map the same areas again. Each area is washed and dried, and a few minutes later the mapping is begun. The stimulaters should be heated to 40° C. and wiped dry before application. Care must be taken to apply the stimulaters with a constant force each time. Moreover, the duration of the stimulus should be very short and constant from trial to trial. The rods should be frequently changed in order to insure constant temperature of the stimulus. Again, a complete map is made.

Cold Spots

Materials. The cold stimuli are also applied by means of temperature cylinders which are cooled to the desired temperature in a beaker of cold water (iced).

Procedure. The temperature cylinders are cooled to a temperature of 15° C. The same procedure is followed as in the mapping of the warm spots.

Treatment of Results

For each sense and for each area, a count of the number of spots is made. Two sets of comparisons are made: (1) for each sense the two areas are compared with respect to density of spots; (2) for a given area the densities of the four types of spots are noted.

We also consider the location and grouping of the spots. Do the pressure spots and pain spots coincide? Are cold and warm spots found in the same squares of the grid? How do the patterns or distributions of the four types of spots compare with each other?

Stability of Spots. The above procedure provides, of course, no data relevant to the stability of spots. If time permits, or as an alternative

procedure to mapping all four senses, the areas may be remapped for any one sense. The reliability or stability of spots is then tested by (1) comparing the total number of spots obtained in the two mappings, (2) noting how many of the individual spots are found on both occasions, (3) counting how many spots found in the first mapping disappear in the second, and (4) ascertaining how many new spots appear in the second mapping.

Change in Intensity of Stimulus. The number of spots found in any one experiment depends on the intensity of the stimulus used. An alternative procedure, then, would be to remap the two areas for pressure spots with a stimulus of different intensity. To obtain an increase in intensity we may use an esthesiometer with a stiffer hair which will exert a greater force per unit diameter. In applying the more intense stimulus, the record of the first mapping should be kept out of the sight of the subject so that he should not be influenced by his previous reports. The following questions should be considered: (1) What is the general effect of increasing the intensity of stimulation? (2) Do the two areas show comparable changes in the density of pressure spots with increase in the level of stimulation?

The other types of sensitive spots may be remapped in a similar fashion with stimuli of greater intensity.

General References

Boring, E. G., Langfeld, H. S., and Weld, H. P., *Psychology: a factual textbook,* New York: John Wiley and Sons, 1935, Chap. 7.

Boring, E. G., Langfeld, H. S., and Weld, H. P., *Foundations of psychology,* New York: John Wiley and Sons, 1948, Chap. 16.

Clark, D., Hughes, J., and Gasser, H. S., Afferent function in the group of nerve fibers of slowest conduction velocity, *Amer. J. Physiol.*, 1935, *114*:69-76.

Gasser, H. S., Conduction in nerves in relation to fiber types, *Proc. Ass. Res. Nerv. Ment. Dis.*, 1934, *15*:35-59.

Jenkins, W. L., Studying the skin senses. In T. G. Andrews (ed.), *Methods of psychology,* New York: John Wiley and Sons, 1948.

Jenkins, W. L., and Stone, L. J., Recent research in cutaneous sensitivity: II. Touch and the neural basis of the skin senses, *Psychol. Bull.*, 1941, *38*:69-91.

Morgan, C. T., *Physiological psychology,* New York: McGraw-Hill Book Co., 1943, Chap. 12.

Nafe, J. P., The pressure, pain and temperature senses. In C. Murchison

(ed.), *A handbook of general experimental psychology*, Worcester: Clark University Press, 1934.

Stone, L. J., and Jenkins, W. L., Recent research in cutaneous sensitivity: I. Pain and temperature, *Psychol. Bull.*, 1940, 37:285-311.

Woodworth, R. S., *Experimental psychology*, New York: Henry Holt and Co., 1938, Chap. 19.

Pressure Sense

Grindley, G. C., The variation of sensory thresholds with rate of application of the stimulus: II. Touch and pain, *Brit. J. Psychol.*, 1936, *27*:189-195.

Guilford, J. P., and Lovewell, E. M., The touch spots and the intensity of the stimulus, *J. Gen. Psychol.*, 1936, *15*:149-159.

Holway, A. H., and Crozier, W. J., Differential sensitivity for somesthetic pressure, *Psychol. Rec.*, 1937, *1*:170-176.

Zigler, M. J., Pressure adaptation time: a function of intensity and extensity, *Amer. J. Psychol.*, 1932, *44*:709-720.

Pain Sense

Dallenbach, K. M., Pain: history and present status, *Amer. J. Psychol.*, 1939, *52*:331-347.

Lewis, T., and Hess, W., Pain derived from the skin and the mechanism of its production, *Clin. Sci.*, 1933, *1*:39-61.

Neff, W. S., and Dallenbach, K. M., The chronaxy of pressure and pain, *Amer. J. Psychol.*, 1936, *48*:632-637.

Stone, L. J., and Dallenbach, K. M., The adaptation of areal pain, *Amer. J. Psychol.*, 1936, *48*:117-125.

Temperature Sense

Bazett, H. C., McGlone, B., and Brocklehurst, R. J., The temperatures in the tissues which accompany temperature sensations, *J. Physiol.*, 1930, *69*:88-112.

Dallenbach, K M., The temperature spots and end organs, *Amer. J. Psychol.*, 1927, *39*:402-427.

Jenkins, W. L., Studies in thermal sensitivity: 1. Adaptation with a series of small circular stimulators, *J. Exper. Psychol.*, 1937, *21*:670-677.

Jenkins, W. L., Studies in thermal sensitivity: 5. The reactions of untrained subjects to simultaneous warm + cold stimulation, *J. Exper. Psychol.*, 1938, *22*:451-461.

Jenkins, W. L., Studies in thermal sensitivity: 16. Further evidence on the effects of stimulus temperature, *J. Exper. Psychol.*, 1941, *29*:413-419.

»»» 4 «««

AUDITION

THE experimental psychology of audition concerns itself with the sensory experience which results from stimulation of the ear. Our lives are full of such experiences: people's voices, the roar of traffic from the street, soothing or exciting music—the infinitely rich variety of sounds which no one could enumerate. These sounds differ in many ways: some are loud and some are soft; whether loud or soft, some sounds are high-pitched, others low-pitched. In addition to these fundamental modes of variation, there are many others. There are rich, full-bodied musical sounds, and there are shrill, grating squeaks and noises. Even two notes identical in pitch and loudness will vary greatly in quality when they come from two different musical instruments. Of course, the sounds we experience also differ in pleasantness and unpleasantness. The problem for experimental psychology is to relate the ways in which auditory experience varies with changes in the physical stimulus.

The advances in the psychology of hearing have closely paralleled the increasing control of the auditory stimulus which physicists and engineers have achieved. Not so long ago, a set of tuning forks and a siren were the most important tools of a psychologist interested in sound. Today, he may utilize the whole array of ingenious devices and measuring instruments which became available with the arrival of the vacuum tube and the radio.

The Auditory Stimulus

In order to relate stimulus and experience, we have to know some of the fundamental physical properties of sound.

Vibration and Simple Harmonic Motion. Many physical objects are set into vibration when energy is applied to them. As an object moves back and forth, it transmits some of its energy to the

surrounding air particles. The air particles move back and forth in the same pattern as the vibrating object. As the air particles collide with each other, the movement spreads away from the vibrating object until the energy imparted to the particles from the object is dissipated. The stimulus for hearing is this kind of vibratory motion.

The simplest kind of vibratory motion is *simple harmonic motion.* The to-and-fro movement of a pendulum is a common example of simple harmonic motion. A moving object executes simple harmonic

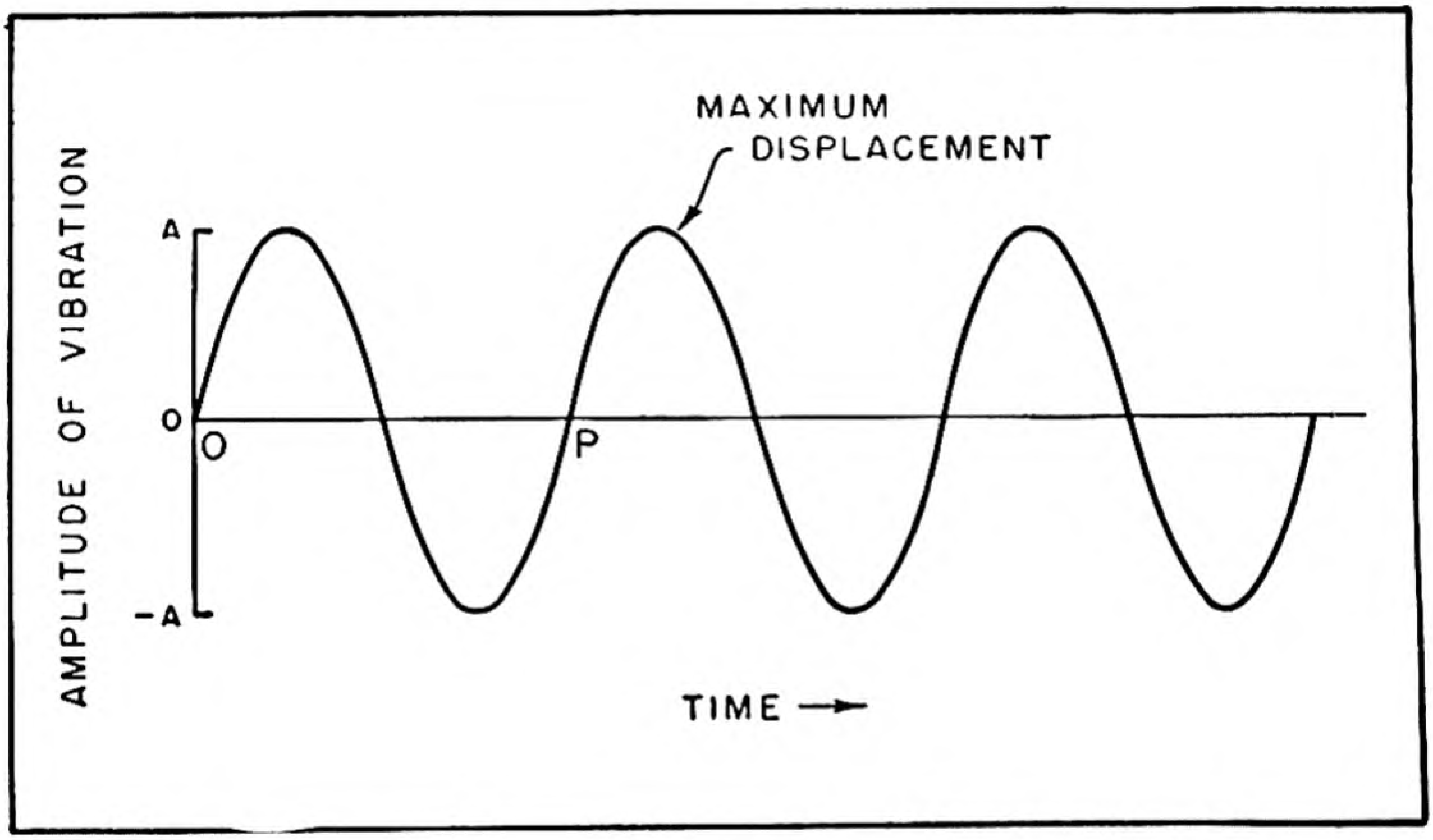


FIG. 5. A Sine Curve, an Example of Simple Harmonic Motion.

motion if the degree of its displacement is proportionate to the force applied to it. When the amount of displacement of such a vibrating object is plotted as a function of time, the *sine curve* shown in Fig. 5 is obtained. The sound emanating from a tuning fork can be described by such a sine curve and is known as a *pure tone.* Along the abscissa of Fig. 5, time is plotted; along the ordinate, the amount of displacement, i.e., the amplitude of vibration. At zero time, the displacement is zero. As time progresses, the displacement rapidly increases in magnitude and reaches a maximum. Once the maximum has been passed, the amplitude decreases back to zero. The displacement of the body then follows the same course, with sign

reversed. When, after going through a maximum and minimum, the displacement has returned to zero, a *cycle* has been completed. The term *cycle* thus refers to the portion of the curve from O to P. The time required for the completion of one cycle is called the *period* of the wave. The number of cycles per second constitutes the *frequency* of the wave. Thus, a sine wave is completely determined by its frequency and its maximum amplitude of displacement.

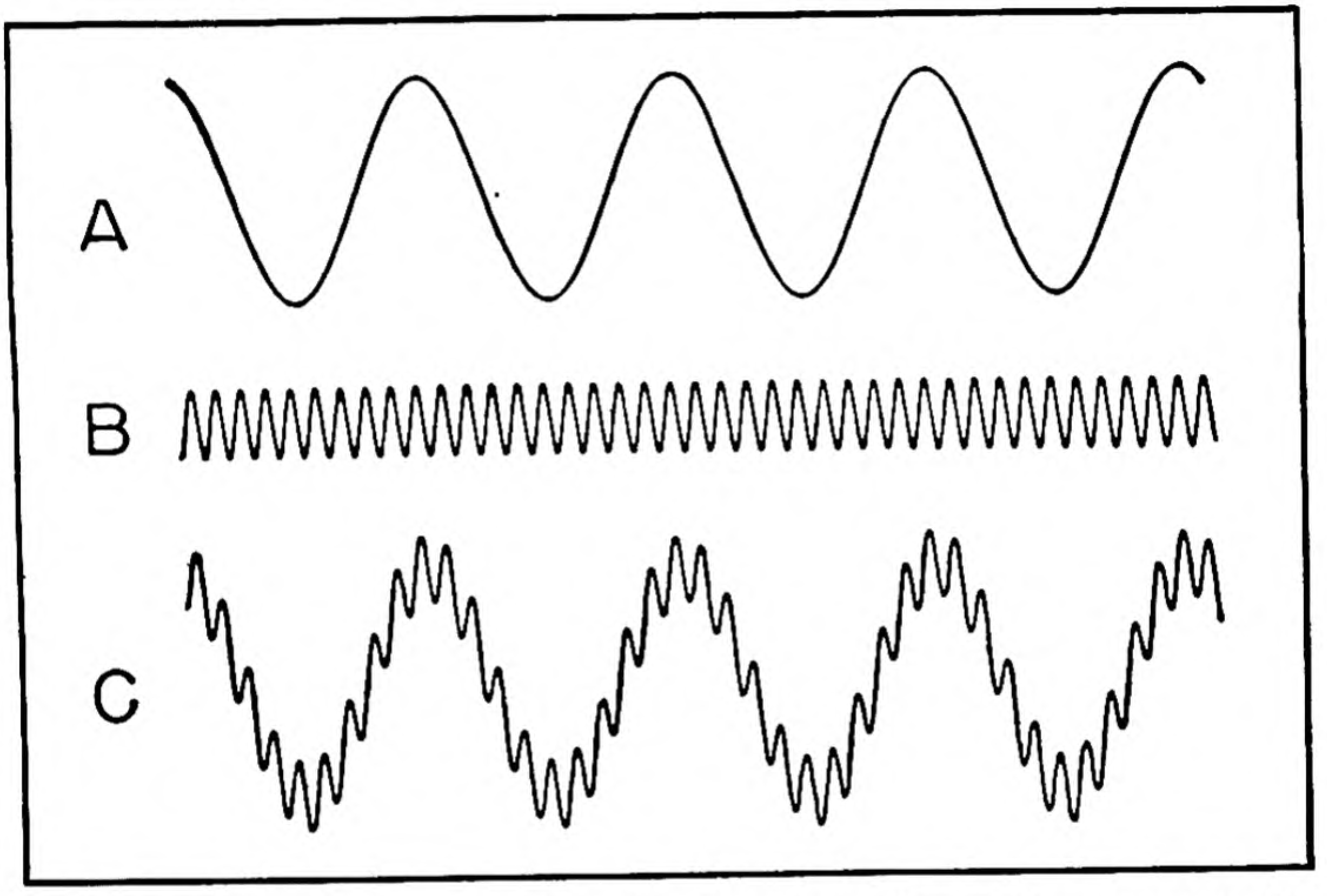


FIG. 6. The Generation of a Complex Tone. Curve C represents the moment-by-moment addition of sine curve A to sine curve B. (After E. G. Boring (ed.), *Psychology for the armed services*, 1945, p. 98, by permission of Infantry Journal, Inc.)

In Fig. 5, the curve leaves the time axis at the origin. Suppose that a wave identical in frequency and intensity is slightly displaced so that it leaves the time axis a short time after the first wave. The two waves would then be out of *phase* with each other. The position on the time axis of one wave with respect to another is thus called *phase*.

Let us return to the vibrations of a tuning fork. The resulting

movement of the air particles creates successive regions of condensation and rarefaction in the medium. These changes in the density of the air give rise to changes in the pressure exerted on any object placed in the vibrating medium. Thus, when the diaphragm of a microphone is placed near a vibrating object (source of sound), the pressure exerted on the diaphragm can be readily measured. It is for this reason that *sound-pressure level* is widely used to measure the intensity of sound.

Only a few of the sounds which we encounter are produced by tuning forks or their electronic equivalents, and, therefore, only few auditory stimuli are simple sine waves. Even a carefully plucked violin string produces a *complex* sound wave. And yet the vibrations of a violin string are extremely simple as compared with the sound waves which we appreciate as speech or the music of a symphony orchestra! With such sounds as speech and music, the complexity of the pattern changes continuously in time. It is to this problem of complexity that we shall now turn.

Consider Fig. 6. There are two sine curves, *A* and *B*, which differ in frequency. The third curve, *C*, represents the moment-by-moment addition of curve *A* to curve *B*. Obviously, curve *C* is not a sine curve, but a complex one. In a similar manner, one could add together any number of sine curves and obtain highly complex curves. In fact, it has been proved that one can take any complex curve and analyze it into a series of sine curves having appropriate amplitude and phase relations. This mathematical law is known as *Fourier's theorem* and is a basic principle of acoustics. It applies only to wave forms which periodically repeat themselves.

The Physical Dimensions of the Auditory Stimulus

There are two basic dimensions along which a pure tone can vary: frequency and intensity. These variations may occur independently of each other. When pure tones are combined, another type of variation occurs: there is a change in the complexity of the stimulus.

Physical Units. It is important to describe the physical stimulus in precise quantitative terms. The frequency of a sound wave is designated in terms of *cycles per second* (c.p.s.). This measure refers, of course, to the number of cycles of a periodic wave that occurs in 1 second. As we have indicated, the *intensity* of the auditory stimu-

lus is most conveniently expressed in terms of sound pressure. Now pressure is defined as force per unit area. The most common unit of pressure in acoustics is *one dyne/cm²*. (One dyne is the force which will increase the velocity of a mass of one gram one centimeter per second every second.) At those frequencies to which the ear is most sensitive, the most intense pressure that the ear can tolerate is more than a million times as large as the weakest pressure that the ear can detect. This tremendous range of sound pressures makes the use of a *logarithmic* scale especially convenient. It is in the nature of a logarithmic scale to compress the upper end of a scale relative to the lower end. Such a compression of the scale is convenient since a change of say, one dyne/cm², in the upper region of intensity does not usually produce nearly as large a sensory effect as it does in the lower region. The logarithmic scale in terms of which sound pressures are expressed is the *decibel* (db) scale.

The number of decibels is defined by the following equation:

$$N_{db} = 20 \log_{10} \frac{p_2}{p_1}$$

where N is the number of decibels, p_1 is a reference pressure, and p_2 is the pressure to which the number of decibels refers. Clearly then, a decibel is a measure of *relative magnitude*. The fraction, p_2/p_1, tells us how many times greater p_2 is than p_1. Then, in order to compress the scale, the logarithm of this fraction is taken. The multiplication by 20 is a matter of convention.

The decibel is a measure of *relative* pressure. It has meaning, therefore, only if the reference pressure is known. To say that a sound stimulus has a pressure of 20 decibels and to let it go at that is meaningless. All that has been said is that p_2 is ten times greater than some (unknown) pressure, p_1. Not until p_1 is specified does such a statement acquire significance. Two reference pressures have been widely used in acoustical research:

Sound-Pressure Level: p_1 here is taken to equal 0.0002 *dyne/cm²*. When the term *sound-pressure level* (SPL) is employed, it is understood that this reference level has been used.

Sensation Level: It is often convenient to take as the reference level not an arbitrary physical magnitude but rather the sound pressure of the stimulus at the absolute threshold. When this reference

pressure is employed, the intensity of the stimulus is expressed in terms of *sensation level*. Thus, to say that the intensity of a sound is 50 db sensation level means that the sound pressure is 50 db above the absolute threshold of the subject for that particular sound.

Auditory Discrimination

Let us now show how auditory experience depends on the physical dimensions of the stimulus.

Determination of Auditory Area

We may best begin with an experimental demonstration. As a source of sound, we shall use an audio-oscillator which generates pure tones (sine waves) whose intensity and frequency we can control. We put our subject in a very quiet room. The oscillator drives an earphone which is fitted snugly over the listener's ear. We set the frequency control dial of the oscillator to 500 c.p.s. We turn the intensity control to a low setting. The subject, who has been instructed to signal when he hears a tone, fails to react. We again present the tone with the intensity slightly increased. On the third such presentation of the tone, the subject signals.

We now begin a new series of presentations. The tone is first presented at an intensity well above the values used in the first series. The subject immediately responds. On successive presentations of the tone, we decrease the intensity until the subject no longer responds. We repeat these two types of series several times and then average those intensities at which the subject's reaction changes (from *no response* to *response* and from *response* to *no response*) This average defines the *absolute threshold* of the subject at 500 c.p.s. The method which we have used in obtaining the threshold is the method of minimal changes (or method of limits).

This procedure is repeated at selected frequencies which range from the lowest frequency to which the subject will respond (about 20 c.p.s.) to the highest (about 20,000 c.p.s.). When the thresholds thus obtained are plotted as a function of frequency, the lower curve in Fig. 7 results. This curve is the sensitivity curve of the human ear.

At each of the frequencies which we have previously used, we now increase the intensity of the tone until the subject reports that he experiences *tickle* or *pain*. When the averages of these intensities

are in turn plotted as a function of frequency, the upper curve in Fig. 7 results: the curve of the *threshold of feeling.*

The sensitivity curve of the ear and the curve of the threshold of

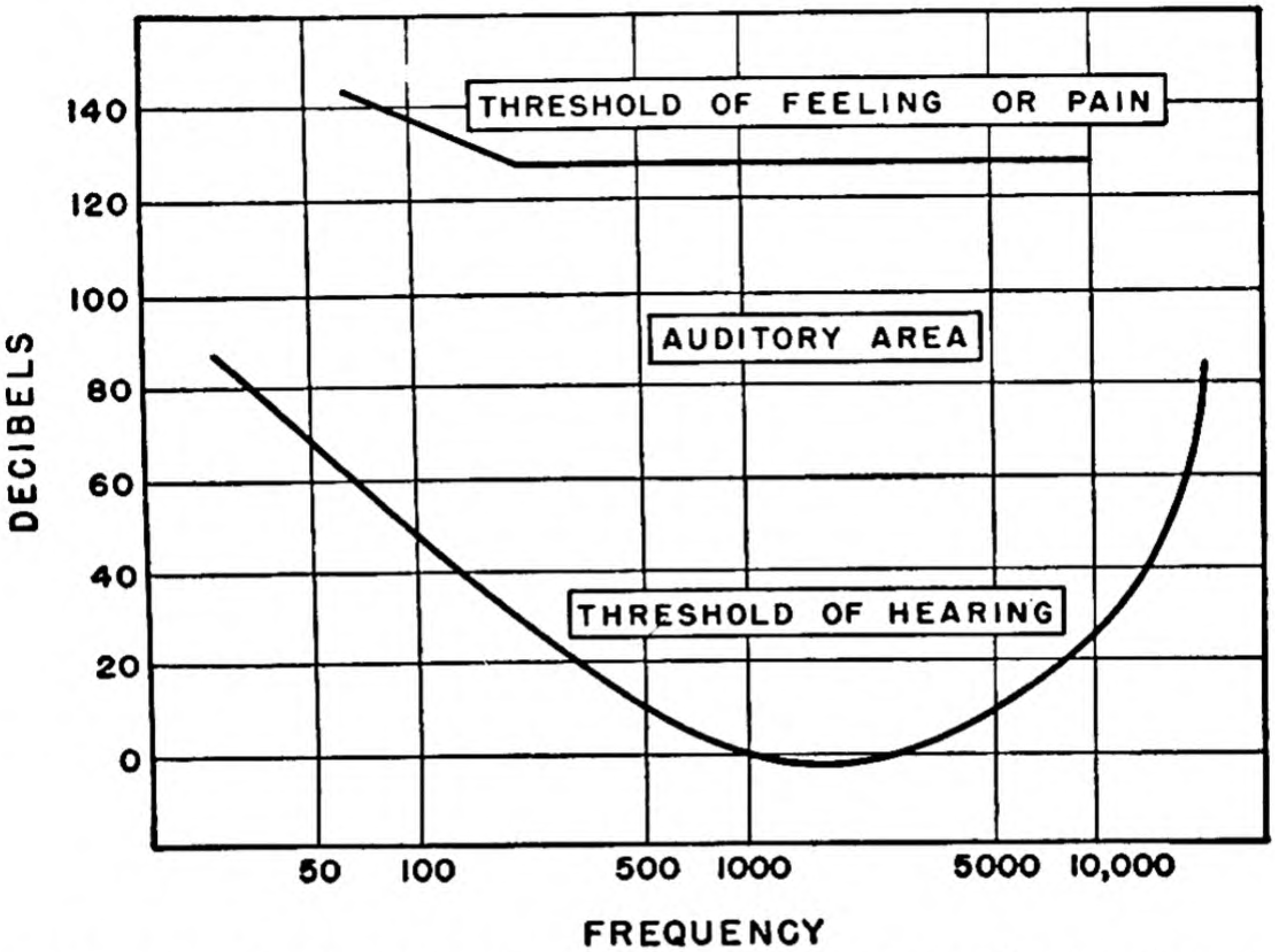


FIG. 7. The Sensitivity Curve of the Human Ear. The lower curve represents the sound pressures needed at the different frequencies to produce a just audible sound. The upper curve represents the sound pressures which result in a feeling of pressure or pain. The area delimited by these two curves is the auditory area. (From E. G. Boring (ed.), *Psychology for the armed services,* 1945, p. 102, by permission of Infantry Journal, Inc.)

feeling form the boundaries of an area within which fall all the stimuli to normal hearing. This area is the *auditory area.*

Differential Sensitivity to Frequency and Intensity

Differential Sensitivity to Frequency. Suppose we present our listener with a 500-cycle tone well above his absolute threshold. Half a second later, we present him with a 502-cycle tone at the

same intensity. The subject has been instructed to indicate whether the second tone is higher or lower in pitch than the first. After presenting this pair of tones, we wait 4 seconds and then again turn on the 500-cycle tone, following it this time with a 496-cycle tone. We continue this procedure, pairing the 500-cycle tone (the stand-

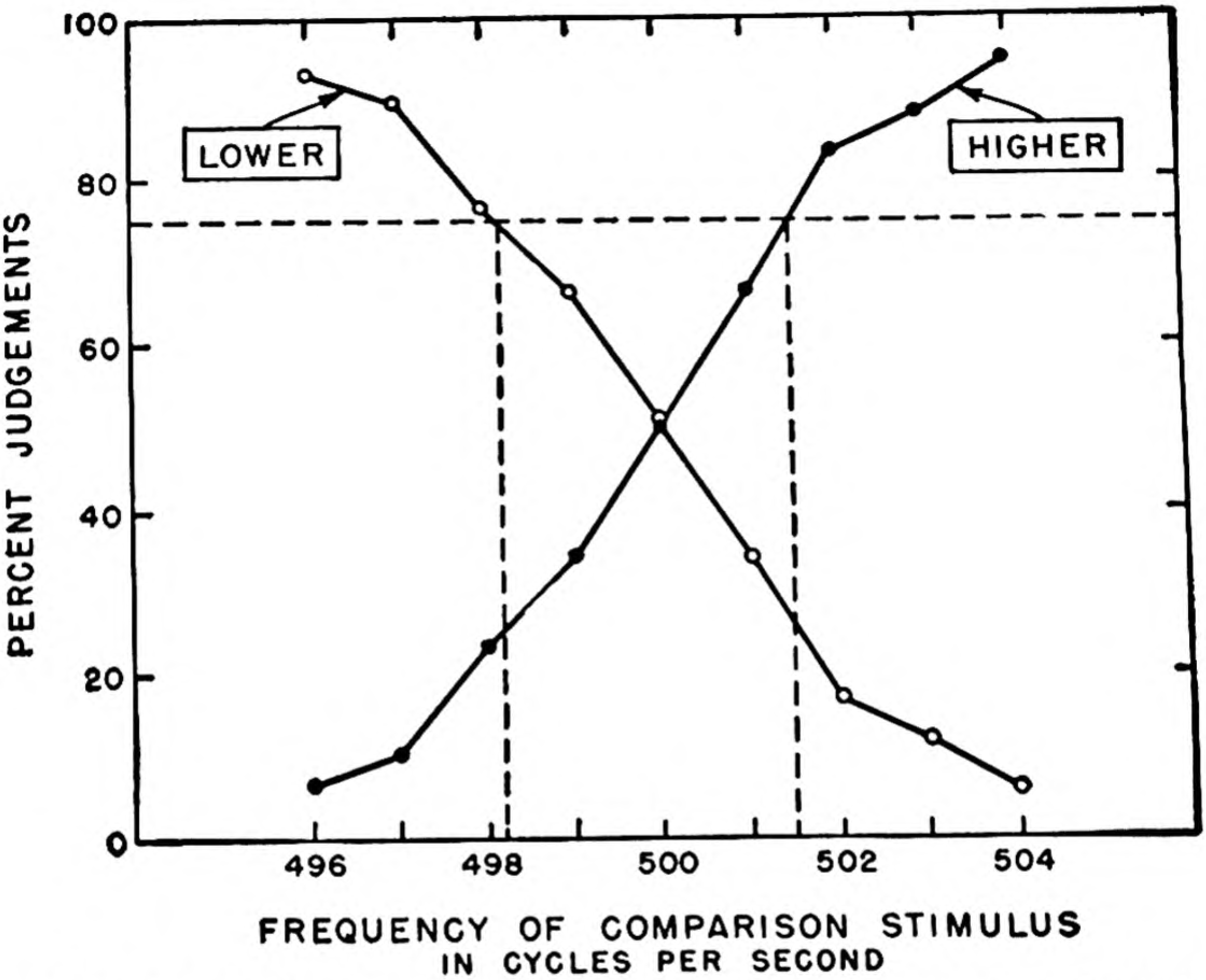


FIG. 8. Distributions of Judgments of *Higher* and *Lower* in an Experiment on Differential Sensitivity to Frequency. The standard stimulus is a 500-cycle tone which is paired with the comparison stimuli indicated on the abscissa.

ard stimulus) with nine different tones (comparison stimuli) covering a range from 496 to 504 c.p.s. Thus, the subject compares the standard stimulus with each of the following comparison stimuli: 496, 497, 498, 499, 500, 501, 502, 503, 504 c.p.s. In short, the method of constant stimulus differences is used. The results are shown in Fig. 8 as psychometric functions. Along the abscissa we have plotted

the frequencies of the comparison stimuli; along the ordinate, percent of judgments in a given category. Our subject has used two categories: *higher* and *lower*. It will be remembered that when two categories of judgment are employed, it is customary to use 75

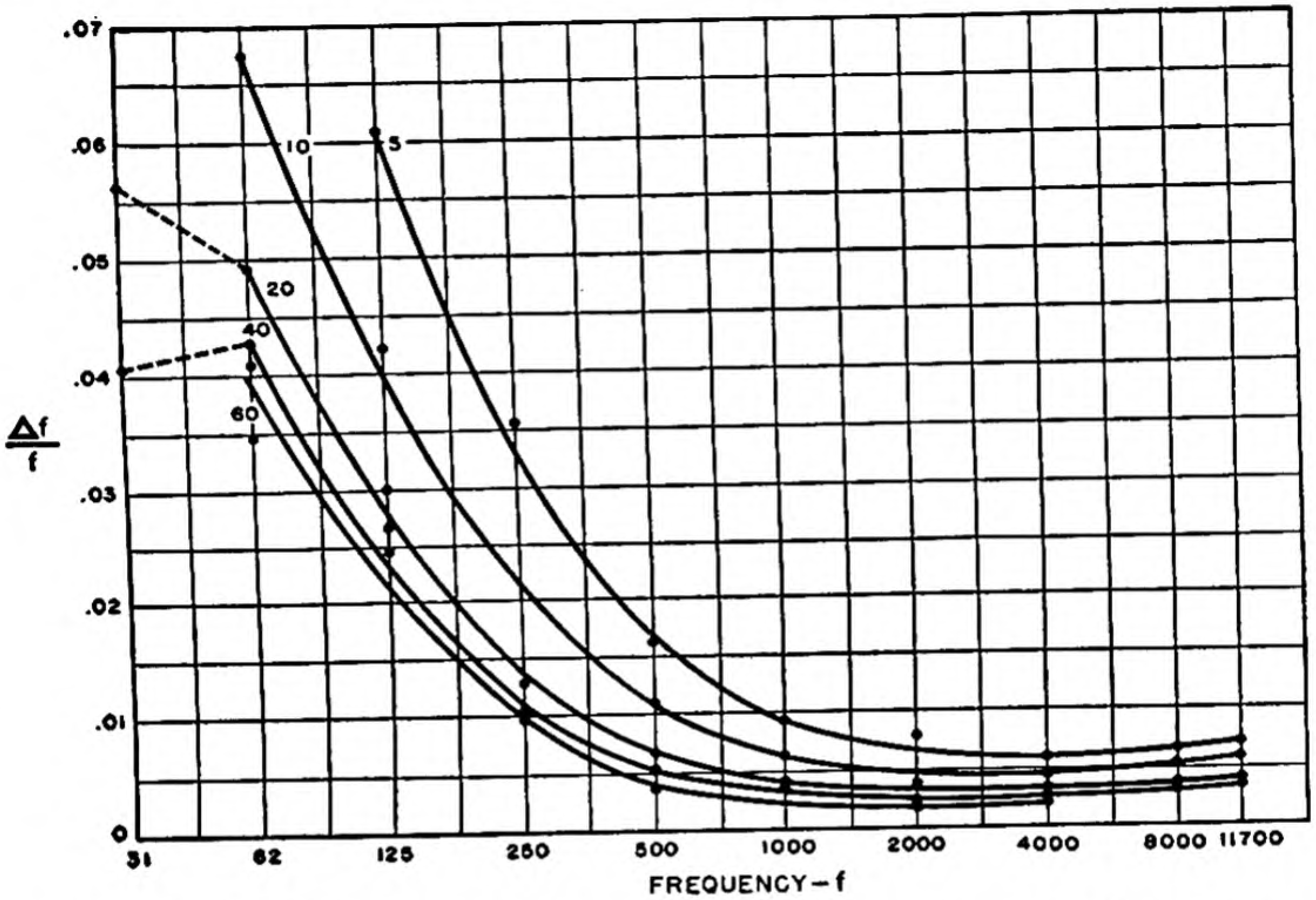


FIG. 9. Variations of $\Delta f/f$ with Frequency, Sensation Level as a Parameter. This family of curves shows how the differential sensitivity to frequency varies with the frequency of the standard tone at various sensation levels. The numbers on the curves refer to sensation levels. (From E. G. Shower and R. Biddulph, Differential pitch sensitivity of the ear, *J. Acous. Soc. Amer.*, 1931, 3.275, by permission of the journal and authors.)

percent correct judgments as a basis for determining the differential threshold. In Fig. 8, the horizontal dotted line represents this 75 percent level; the vertical dotted lines have, then, been drawn to find the comparison stimuli corresponding to this percentage.

We have now measured the differential threshold for frequency with a standard of 500 c.p.s. This value applies only to this frequency at the intensity which we have used. We are interested in

the functional relation between the value of the differential threshold and the frequency of the standard stimulus. Since the relation between the differential threshold and frequency also depends on the intensity at which the determinations are made, we need a family of curves to represent the relations involved. Fig. 9 shows such a family of curves. On the abscissa is plotted frequency in c.p.s. The ordinate represents values of $\triangle f/f$, i.e., *relative* differential thresholds. Each of the curves was obtained at a different intensity of the stimuli. Therefore, the plot represents $\triangle f/f$ as a function of frequency with intensity as the parameter. A parameter is, of course, a variable which is held constant for a given series of determinations but varied from series to series. The plot illustrates the joint dependence of the differential threshold on frequency and intensity.

Differential Sensitivity to Intensity. We may use the same experimental procedure that we described for the measurement of $\triangle f/f$ for the determination of $\triangle I/I$. This time, of course, the method of constant stimulus differences is applied to intensity changes, and frequency is the parameter. This set of relationships is represented in Fig. 10.

Attributes of Auditory Experience

As we have seen, listeners are able to make consistent and reliable responses when instructed to judge tones as *higher* or *lower* and as *louder* or *softer*. The listener's ability to make consistent and reliable judgments with respect to these categories establishes the concepts of *pitch* and *loudness*. Just as frequency and intensity are dimensions of the physical stimulus, so pitch and loudness are dimensions of auditory experience. It is of utmost importance to keep in mind that pitch is not the same as frequency, and loudness not the same as intensity. Pitch and loudness are aspects of auditory experience which are inferred from the subject's responses to tonal stimulation. We shall now describe the principal functional relationships between the physical dimensions of the stimulus and the attributes of auditory experience.

Pitch as a Function of Frequency. Changes in pitch are largely, though not exclusively, a function of frequency. In general, as the frequency of a tone is increased, the pitch rises. The rela-

tionship between pitch and frequency may be determined in a number of ways. One particular function was obtained by requiring subjects to adjust the frequency of a comparison stimulus until it sounded just half as high in pitch as the standard stimulus. By selecting standard stimuli over the audible range, it was then pos-

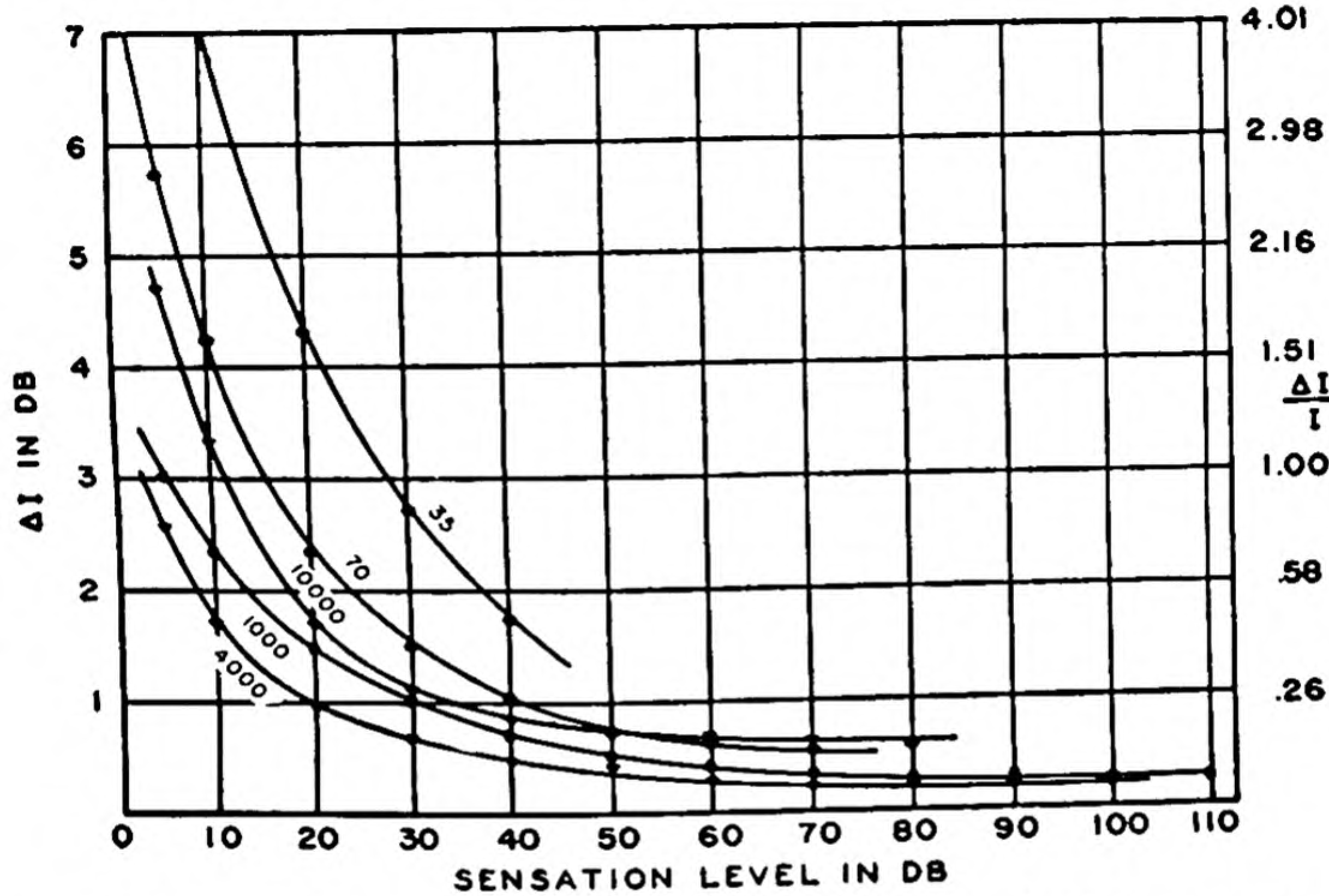


FIG. 10. $\Delta I/I$ as a Function of Sensation Level with Frequency as a Parameter. This family of curves shows how differential sensitivity to intensity depends on sensation level at different frequencies. The size of the differential threshold (ΔI) is plotted on the left and the size of the relative differential threshold ($\Delta I/I$) is plotted on the right. The numbers on the curves refer to frequency. (Reproduced by permission from *Hearing: its psychology and physiology* by S. S. Stevens and H. Davis, p. 138, published by John Wiley & Sons, Inc., 1938.)

sible to determine the frequency steps required to obtain equal increments in pitch. The unit of pitch was obtained by arbitrarily calling the pitch of a 1000 c.p.s. tone, 1000 mels. The pitch of a tone sounding just half as high would then be 500 mels. As the function in Fig. 11 shows, the frequency of a tone of 500 mels is not 500 c.p.s., but is 400 c.p.s. Clearly, the relationship is by no means linear; for

equal increments in frequency, one does not obtain equal increments in pitch.

There is a striking phenomenon which points up the fact that

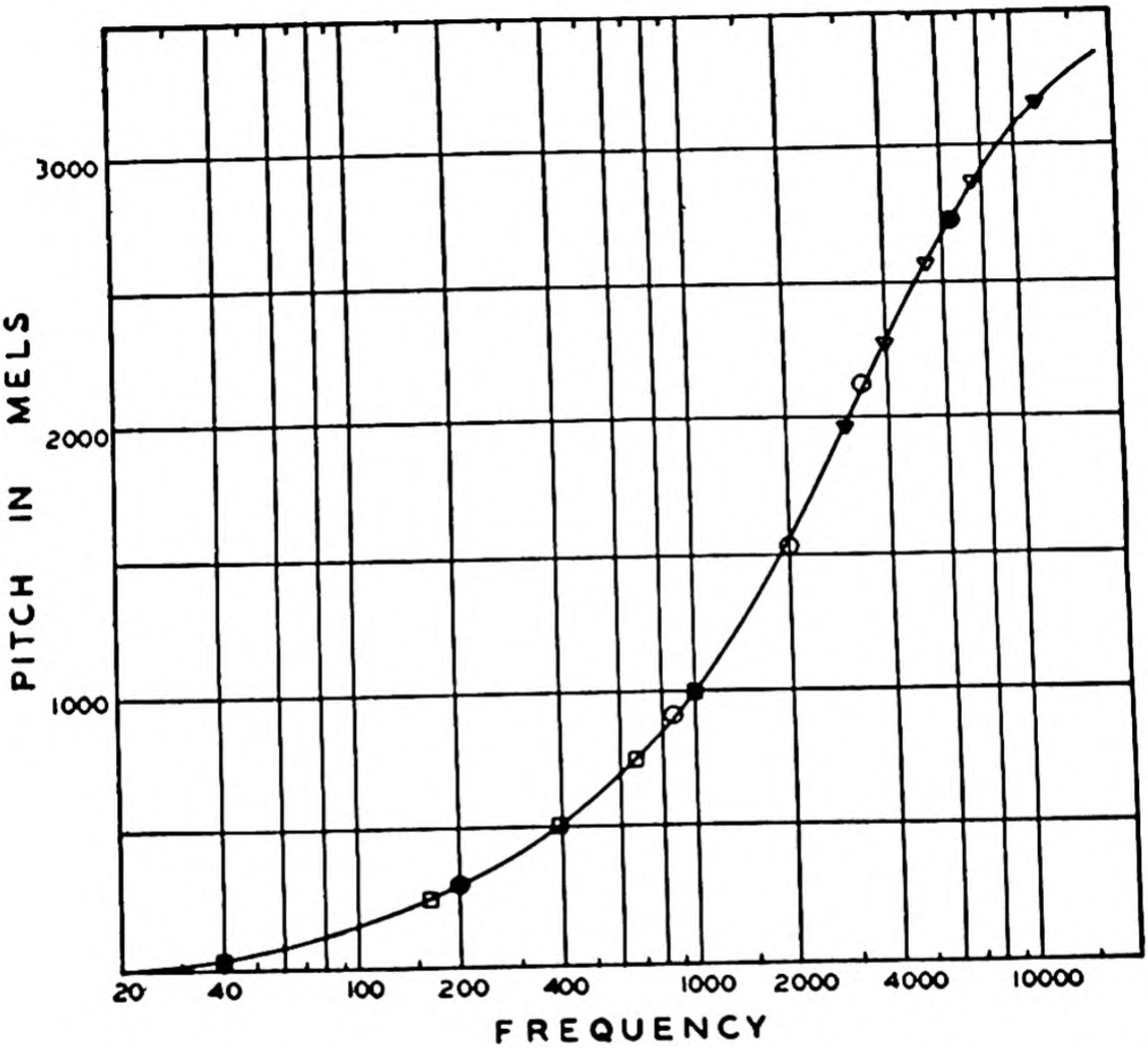


FIG. 11. The Pitch Function. This curve shows how pitch, measured in subjective units (mels), grows with increases in frequency. Note that the relation between frequency and perceived pitch is not linear. (From S. S. Stevens and J. Volkmann, The relation of pitch to frequency: a revised scale, *Amer. J. Psychol.*, 1940, 53:336, by permission of the journal and the authors.)

pitch is not the same as frequency. If a subject is required to match the pitch of a tone heard in one ear with the pitch of a tone heard a moment later in the other ear, the pitch match may be achieved

even though there is a fairly large difference in frequency between the stimuli. This phenomenon is known as *diplacusis.*

Pitch as a Function of Intensity. Although the pitch of pure tones is a major function of frequency, investigators have found that it also varies to some extent with intensity. The pitch of low tones is especially affected by intensity. Tones of low frequency (less than about 500 c.p.s.) sound lower in pitch as they become louder. There has also been some indication that tones of high frequency (more than about 3000 c.p.s.) sound higher in pitch as they become louder. The evidence for complex tones (musical tones) is that they remain constant in pitch as intensity changes.

Loudness as a Function of Intensity. Loudness is a major function of intensity. As the intensity of a tone is increased from its threshold value, the loudness, at first, increases only a little. As the intensity is increased further, the loudness increases more and more. Fig. 12 shows loudness in subjective units (sones) plotted as a function of intensity. The loudness function was obtained by the same experimental procedure as the pitch function shown in Fig. 11.

Loudness as a Function of Frequency. Loudness also varies to a considerable extent with frequency. The sensitivity curve of the ear shows that for minimal audibility, different intensities are required at different frequencies. In order to maintain a somewhat greater than minimal (threshold) loudness, different intensities are also required at different frequencies. In order to maintain a very high degree of loudness, however, the intensity may be held virtually constant as frequency varies. Fig. 13 shows a set of *equal-loudness contours.* We obtain an equal-loudness contour when the intensity is adjusted at each frequency so as to maintain a constant level of loudness. Thus, when a 1000-cycle tone has a sensation level of 40 db, a 100-cycle tone must be only 25 db above threshold, to sound equally loud. It may be noted, however, that the physical intensity of the 100-cycle tone must be greater than that of a 1000-cycle tone.

Other Attributes. Pure tones have attributes other than pitch and loudness. For example, the attribute of *volume* refers to the aspect of largeness or smallness of tones. Low tones and loud tones are large, while high tones and soft tones are small. Of course, musical tones and other complex tones have many other psychological dimensions.

Although pitch is best defined by using pure tones as stimuli, sounds of considerable complexity still have a definite, characteristic pitch. The pitch of a violin note may be readily identified even though there are many overtones in this stimulus. On the other

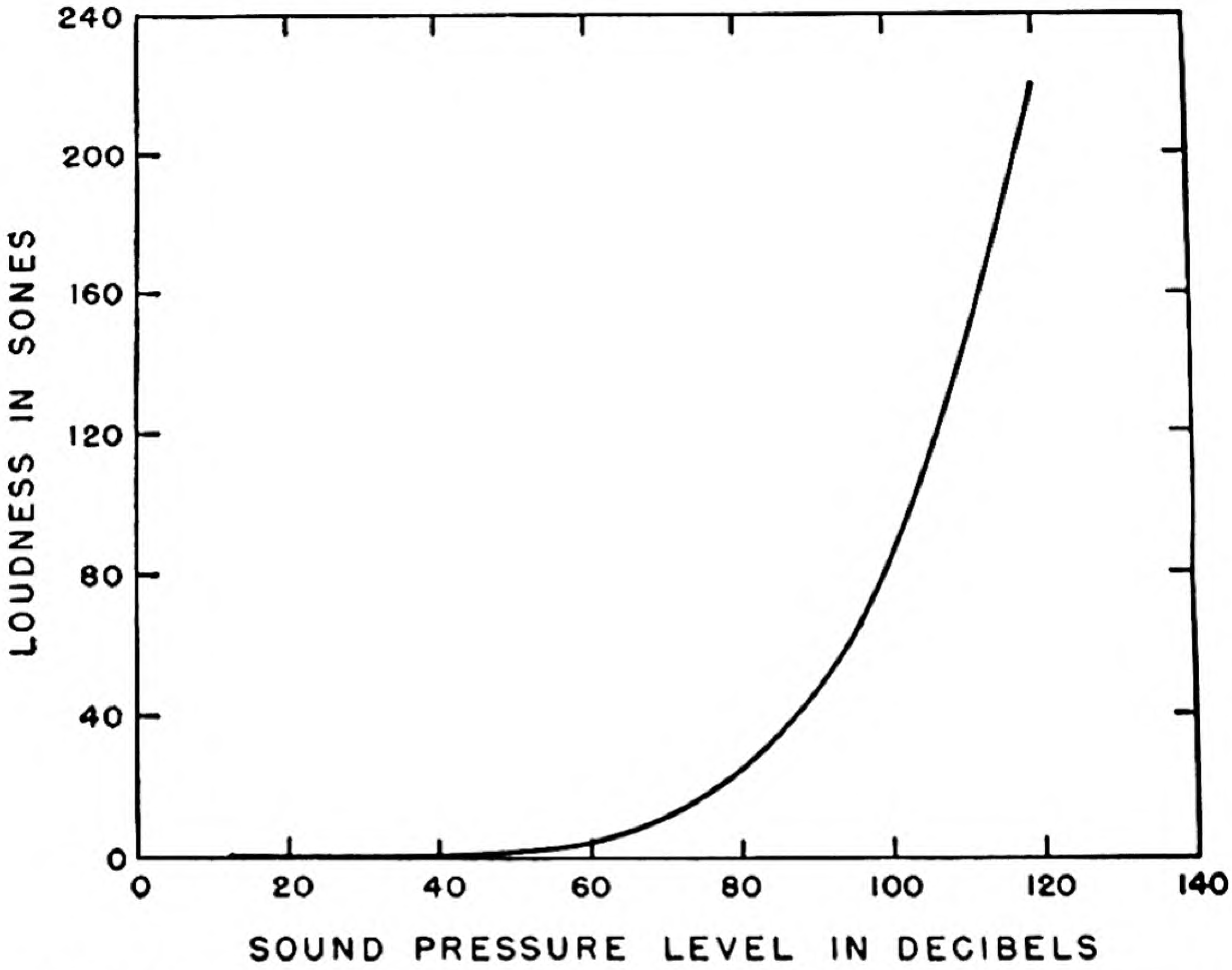


FIG. 12. The Loudness Function. The curve shows how loudness measured in subjective units (sones) grows with increases in sound pressure.

hand, there are complex stimuli whose pitch is evasive. We may say that a thumping noise sounds lower in pitch than a sharp click, but we have difficulty in placing it with much success on the pitch scale. Noises have been called *atonal* as distinguished from musical tones which have tonality, i.e., distinct pitch. This atonality of noises is not surprising if we recall that a noise has many component frequencies which do not bear a simple numerical relation to each other.

The Dependence of Attributes on Both Frequency and Intensity. The functional relationships between each attribute and the physical dimensions of the stimulus make it clear that no single attribute is uniquely determined by a single physical dimension.

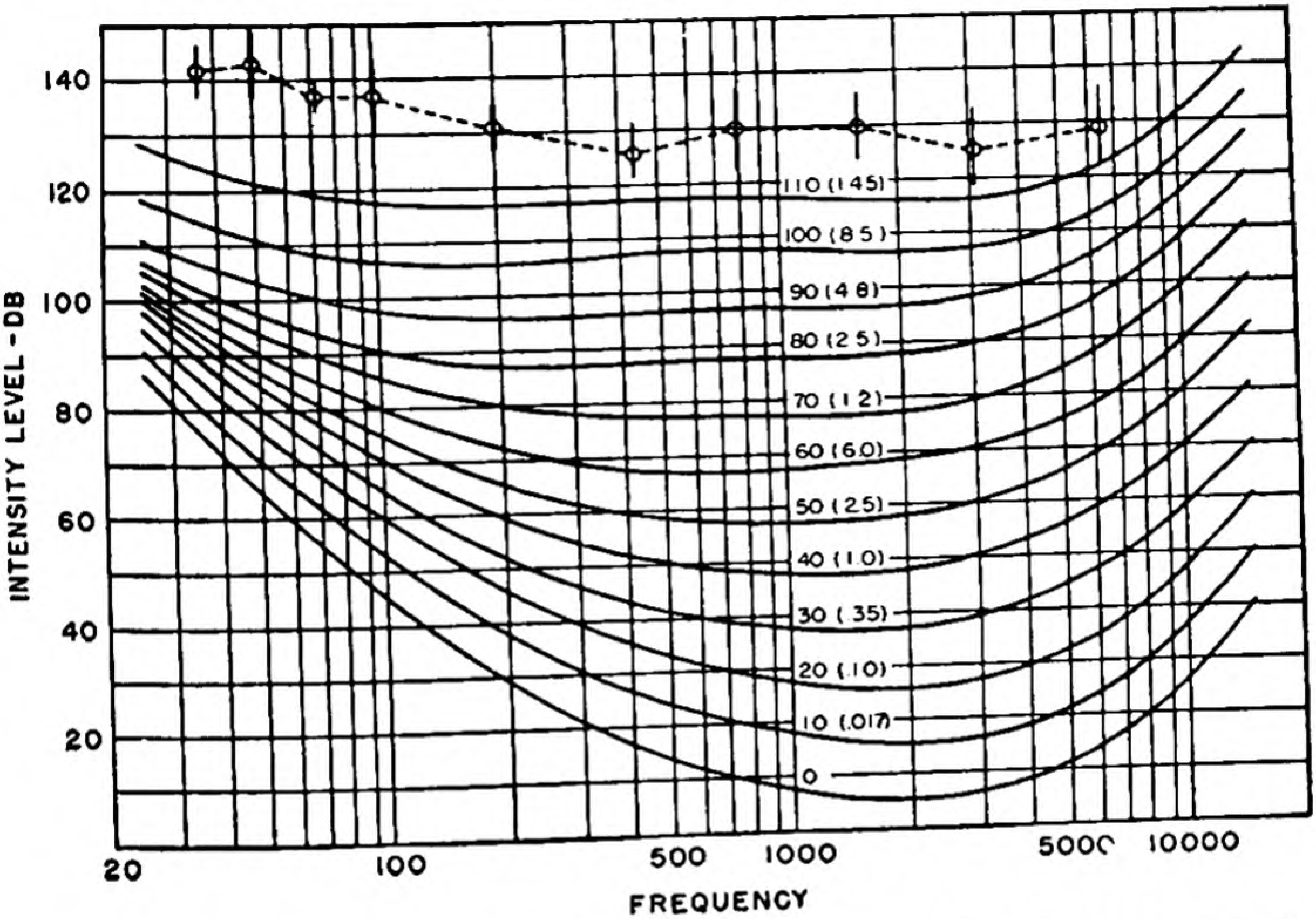


FIG. 13. Equal-Loudness Contours. An equal-loudness contour is obtained when the intensity is adjusted at each frequency so as to maintain a constant level of loudness. The contour marked 0 DB is the sensitivity curve of the ear. The dotted curve refers to the threshold of feeling. The other curves represent different loudness levels. (Reproduced by permission from *Hearing: its psychology and physiology* by S. S. Stevens and H. Davis, p. 124, published by John Wiley & Sons, Inc., 1938.)

Both loudness and pitch are joint functions of frequency and intensity.

PHYSIOLOGICAL BASIS OF PITCH AND LOUDNESS

The great sensitivity of the ear to differences in frequency and intensity at once raises the question of the physiological basis of

this remarkable discriminatory power. How do small changes in the auditory stimulus lead to differential response? What are the dimensions or modes of variation in the auditory nerve which provide the basis for correct discrimination?

First let us consider differential response to changes in frequency. It was once believed that the frequency of nerve impulses duplicated the frequency of the stimulating sound wave. Such an explanation became untenable when it was discovered that a nerve fiber cannot respond with anything near the highest audible frequencies. The refractory periods, both absolute and relative, prevent the fiber from responding more than a few hundred times per second.

The alternative view, long held by many investigators, maintains that stimuli of different frequencies activate different regions of the basilar membrane in the inner ear, and thereby stimulate different nerve fibers. This *place theory* of pitch perception is supported by a considerable array of evidence from anatomical structure, pathological conditions, and electrical responses of the auditory system.

One particularly conclusive experiment utilized the technique of isolating the response to tonal stimulation of a single nerve element. After placing microelectrodes upon a single element, the investigators determined the absolute threshold of that element to auditory stimuli. They found that a given element had its minimum threshold for a particular frequency and that on both sides of this frequency, the thresholds rapidly increased. Furthermore, different elements have their minimal thresholds at different frequencies. The range over which a given element can respond is surprisingly restricted. The auditory areas of selected single nerve elements are shown in Fig. 14.

We do not as yet know how the ear as a mechanical device can analyze the sound wave so that different regions of the basilar membrane are activated by different frequencies. It is quite probable, however, that frequency discrimination is mediated by the stimulation of different nerve fibers.

Which fiber responds, then, seems to be the critical factor in the perception of pitch. But how are discriminations of intensity mediated? Research in neurophysiology has pointed to two mechanisms which probably supplement each other in determining loudness. In the first place, as the intensity of the stimulus is increased, the

thresholds of more and more single fibers are reached. An intense stimulus throws more fibers into action than a weak one. This particular explanation of loudness has been termed the *multiple-fiber hypothesis.* There is also another way in which the nerve responds to an increase in the intensity of the stimulus, viz., by an increase in

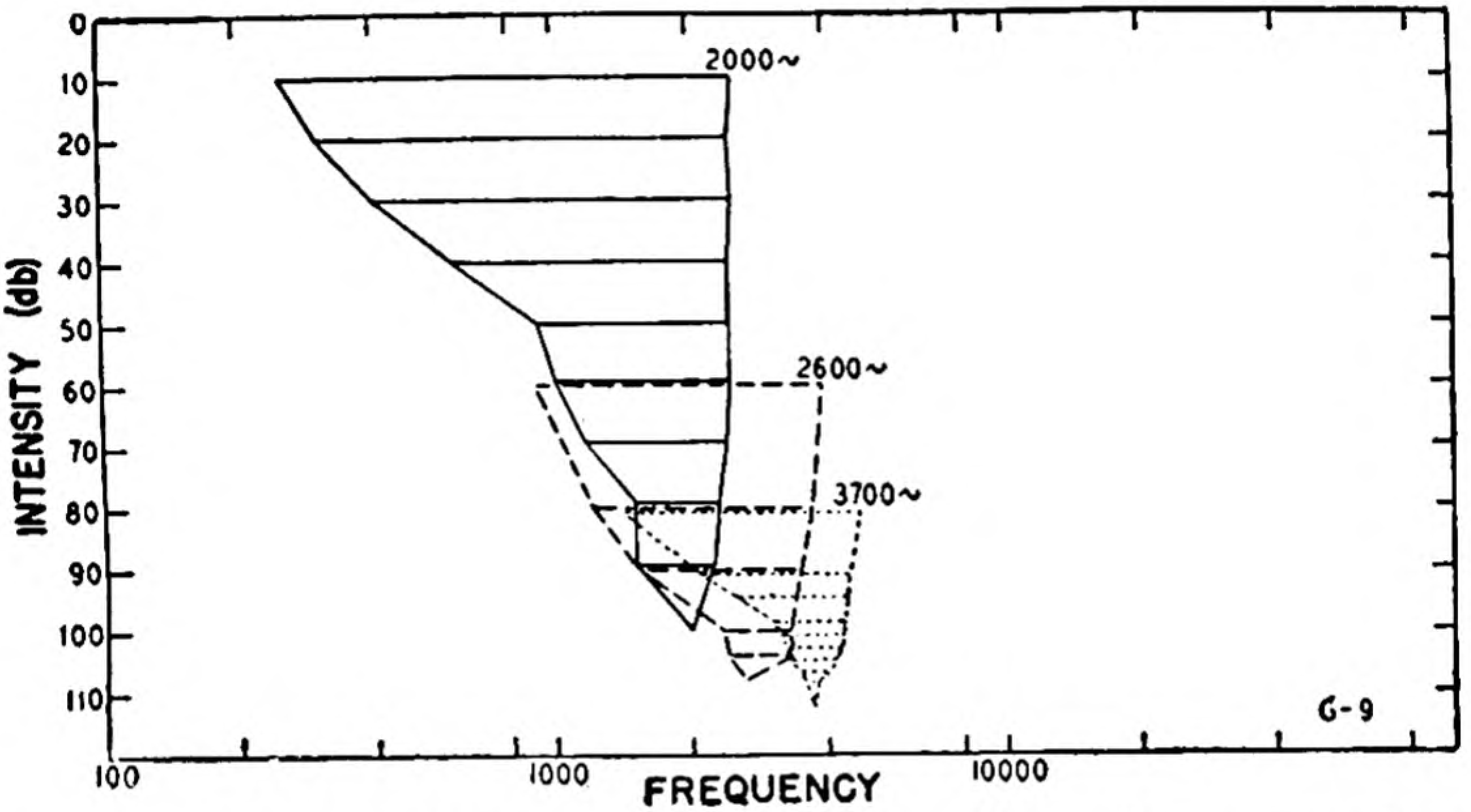


FIG. 14. The Sensitivity of Single Auditory Nerve Elements to Different Frequencies of Sound. The intensities (in db) required for activation of the single element at various frequencies are shown. Each element responds only over a limited, characteristic range of frequencies. The elements are identified by the frequencies to which they are most responsive. (By permission from *Physiological psychology* by C. T. Morgan, p. 235, copyrighted 1943, McGraw-Hill Book Co., Inc. After R. Galambos and H. Davis, The response of single auditory-nerve fibers in acoustic stimulation, *J. Neurophysiol.*, 1943, 6:45, Charles C. Thomas, publisher, Springfield, Illinois. By permission of the journal and authors.)

the frequency of impulses in each individual fiber. This view of sensory intensity is known as the *frequency hypothesis.* These two mechanisms—an increase in number of fibers excited and an increase in the frequency of impulses—undoubtedly work together. Each process by itself, or both together, serve to increase the total number of impulses per unit of time in the nerve. It is this single variable, total number of impulses per unit time, upon which loudness depends.

Beats, Difference Tones, and Masking

In this section, we shall concern ourselves with the auditory phenomena resulting from the stimulation of the ear by *two* pure tones.

Beats. Let us take pure tones of equal intensity somewhat above the absolute threshold, one at 1200 c.p.s., the other at 1700 c.p.s. The subject reports that he hears two tones distinctly differing in pitch. Let us keep the lower tone fixed at 1200 c.p.s. but gradually decrease the frequency of the higher one. When the two tones are about 150 c.p.s. apart, the subject still reports hearing two pitches but now describes them as "rough." When we further decrease the frequency difference between the two tones to about 50 c.p.s., an intermittent series of pulses is heard, and when the tones are about 10 c.p.s. apart, a tone of a single pitch is heard which waxes and wanes in loudness. It is this waxing and waning of loudness that is called *beats.*

The rise and fall in loudness is due to the continually changing phase relation between the two stimuli. The two stimuli are acting upon a common region of the basilar membrane in the inner ear, which contains the receptor cells for hearing. At one moment of time the two stimuli reinforce each other and thus make a more intense stimulus than either alone. At this time the loudness is maximal. As the phase-difference between the stimuli shifts, they come to cancel each other in their effect on the basilar membrane. As a result, the perceived loudness falls.

There are two physical stimuli present, yet the subject hears only one pitch. Clearly, the capacity of the ear to analyze tonal stimuli is limited.

Difference Tones. Let us return to our original stimuli: two pure tones of equal intensity, one at 1200 c.p.s., one at 1700 c.p.s. These tones are first presented at an intensity just above the absolute threshold. The subject again reports two distinct pitches. We now increase the intensity of both tones to an equal extent: to, say, 80 db sensation level. The subject now reports that he hears not only the two original tones but also a third tone whose pitch matches that of a 500-cycle tone. This third tone is the *difference tone.* Its pitch corresponds to that of a tone whose frequency is equal to the difference between the two primary stimuli. Difference tones

may be ascribed in part to the fact that as the stimuli are transmitted through the bony chain of the middle ear to the receptors in the inner ear, distortion of the stimulus pattern occurs. The ear

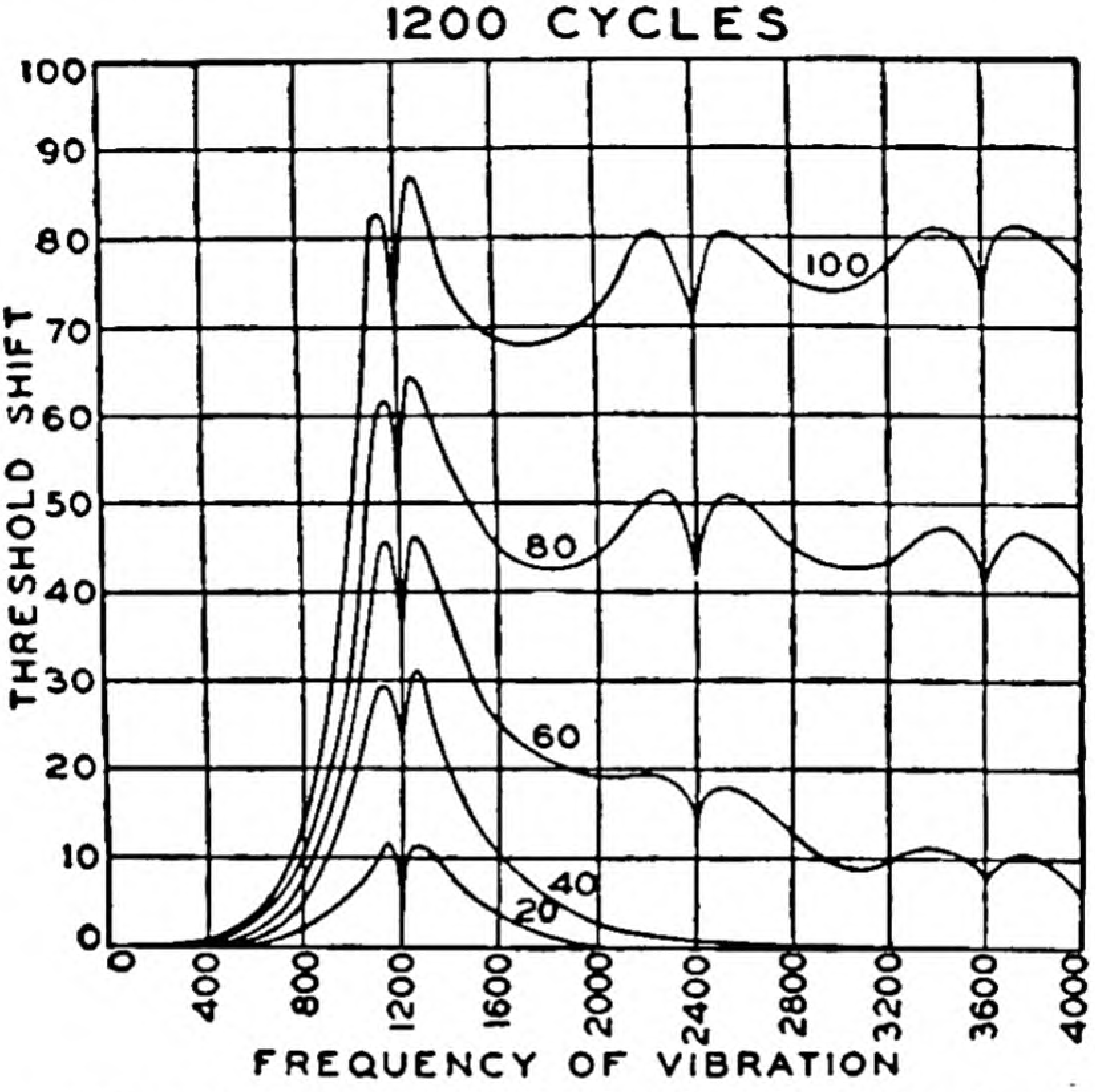


Fig. 15. Masking. This family of curves shows the extent to which a 1200-cycle tone at different sensation levels (parameters) masks tones of other frequencies. (From H. Fletcher, *Speech and hearing*, 1929, p. 169, by permission of D. Van Nostrand Company, Inc., courtesy of Bell Telephone Laboratories.)

as a mechanical device does not convey the vibrations faithfully to the receptors. When two tones pass through such a distorting system, an additional tone is generated. It is this additional tone, present in the inner ear as a physical vibration, which is heard as the difference tone. Note that the mechanism which gives rise to difference tones is very different from that responsible for beats.

Masking. Working with the same two tones, we keep the

intensity of the 1200-cycle tone constant at a sensation level of 80 db. We gradually decrease the intensity of the 1700-cycle tone from a sensation level of 80 db to about 50 db. The subject now reports that he can no longer hear the higher tone. The 1200-cycle tone has *masked* the higher tone. If we now turn off the 1200-cycle tone, the 1700-cycle tone reappears. After all, this tone is 50 db above the subject's threshold measured in quiet. This shift in the threshold of a tone due to the presence of another tone is called *masking*. Since the threshold of the 1700-cycle tone was shifted 50 db, the *amount of masking* in this case was 50 db. In general, the amount of masking will depend on the frequency and intensity relations between the *masking tone* and the *masked tone*. As Fig. 15 shows, a tone masks tones above it in frequency more readily than tones that are below it.

Aural Harmonics

The distortion responsible for the introduction of a difference tone manifests itself in another way. When only one pure tone falls on the ear, its nonlinear transmission introduces *harmonics*. Harmonics are tones whose frequencies are integral multiples of the fundamental, or stimulus, tone. Thus, if a fairly loud tone of 500 c.p.s. is the stimulus, its harmonics will be 1000 c.p.s., 1500 c.p.s., etc. These harmonics, as a result of the distortion in the ear, are present as actual physical vibrations and are, therefore, *aural harmonics*. They may be heard with some practice as tones whose pitch depends upon the frequencies of the harmonics. Loud tones, then, are never heard as pure; there are always aural harmonics in addition to the fundamental tone.

Localization of Sounds

We do not ordinarily hear just "pitches" and "loudnesses"; rather, we hear sounds emanating from objects in our environment. Sounds help to identify objects and events and localize them in space. What aspects of sound make it possible for the listener to identify and to localize objects? The answer is that man can localize sound because he has *two* ears. The position of the two ears on opposite sides of the head creates for the listener differences among sounds of different spatial origins. A sound coming from the listener's right-hand

side will arrive at the right ear first. Thus, there is a *difference in time of arrival* depending on the side from which the sound comes. When the sound comes from the right, there is also a *difference in intensity* in favor of the right ear. This difference is due to the fact that the head casts a *sound shadow* which obstructs the propagation of high, but not low, frequency components of the sound.

Even if a pure tone is continuous, and of the same intensity in each ear, localization is possible. There is still a difference in time of arrival at the two ears for each successive sound wave. In this case, we speak of a difference in *phase*.

It is on these differences among sounds of different location that localization principally depends. Considering the location of the two ears, it is not surprising that localization is most accurate when we have to distinguish between right and left. Sounds originating directly in front of the listener are frequently confused with sounds which come from behind. In this case, the sound source is virtually equidistant from the two ears and the necessary differences are lacking.

Of course, the listener's ability to localize a sound also depends on how familiar he is with it. A loud fire siren must be close, a faint wail is judged as distant. Although the relation between distance and loudness is learned, the listener makes his judgment quickly and confidently.

Auditory Fatigue

Until recently, the ear has been considered to be virtually unfatigable. Even after long exposure to sounds of ordinary intensity, the absolute sensitivity of the ear is little changed. In contrast to the eye, nose, tongue, and skin, whose sense organs continually undergo extensive adaptation, the ear maintains its sensitivity to a remarkable degree. This stability of the sensitivity of the ear is undoubtedly due to the mode of action of the auditory stimulus. The alternating pressures constitute a continually changing stimulus to the auditory receptors. Nevertheless, as modern work has shown, the ear is subject to fatigue. This fatigue is not sufficiently marked to be of importance unless the ear is subjected to severe acoustic stresses.

The most common intense stimuli that the ear has to tolerate are

the noises encountered in the presence of high-powered machinery, as in airplanes, or in the notoriously loud boiler factories. When these noises are brought into the laboratory so that we can control the time of exposure and their intensity, dramatic decreases in sensitivity can be induced. Exposure in the laboratory has the advantage

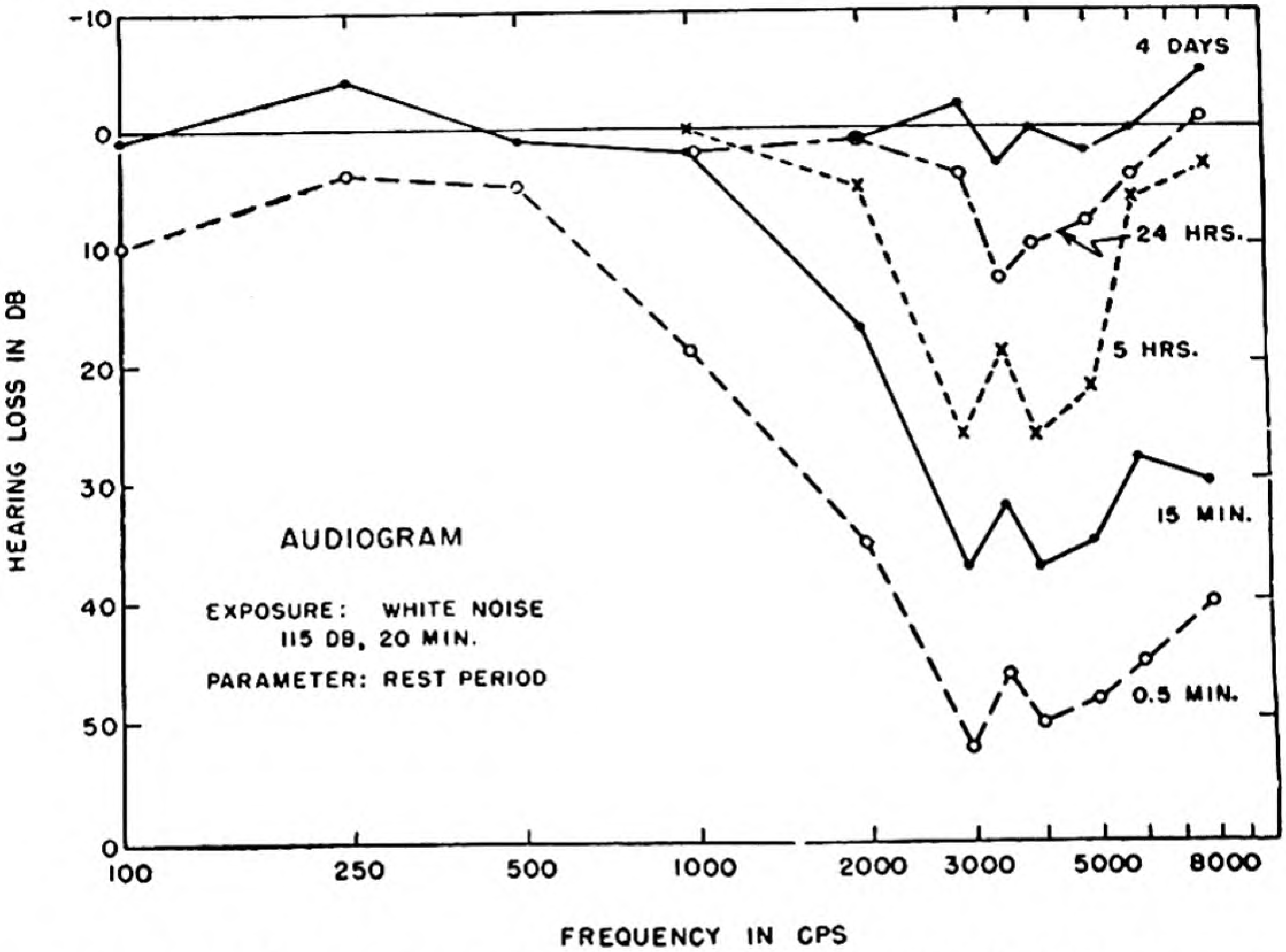


FIG. 16. Changes in Auditory Sensitivity as a Result of Exposure to Intense Noise. Each of the curves represents an audiogram taken at a given time interval after exposure. (From J. P. Egan, unpublished data.)

of making possible determinations of auditory sensitivity before as well as after exposure.

There is an almost unlimited variety of noises that can be produced by emphasizing the intensities of certain frequencies rather than others, but one particular noise is especially suitable for the investigation of auditory fatigue. All audible frequencies are present in this noise, and all of the frequencies are equally intense. This noise has been called *thermal noise* because it may be generated from the thermal agitation of molecules in a resistor. It has also

been called *white noise* by analogy to white sunlight, which contains all visible wave lengths in about the same amount. Let us turn to some experiments on auditory fatigue in which the ear was exposed to this particular noise.

To measure the effect of exposure to intense stimuli, we first determine in the quiet the sensitivity curve of the ear. The ear is then exposed to a loud stimulus, say, a white noise whose overall intensity is greater than 100 db. Immediately following the exposure the sensitivity of the ear is redetermined. The shifts in the absolute thresholds expressed in decibels are taken as measures of auditory fatigue. The temporal course of recovery from fatigue can be traced by repeated determinations of the absolute threshold at various time intervals after exposure.

Fig. 16 shows the amount of fatigue induced by exposure to a white noise (115 db overall). The horizontal zero line represents the normal sensitivity of the ear. The other curves define the temporary hearing losses in decibels for various rest periods after exposure. The curve, taken 0.5 minutes after exposure, shows the striking impairment in sensitivity immediately following exposure to intense noise. The curves taken after longer rest periods gradually approach the normal sensitivity curve, and there is complete recovery after about four days. Noteworthy is the fact that the hearing losses are most pronounced at high frequencies. Just why the ear is most vulnerable in the region of high frequencies is not known.

Speech, Hearing, and Communication

The experimental psychology of audition extends far beyond the study of pure tones. It is through the use of pure tones that we are best able to investigate the basic dimensions of auditory experience. But the sounds which are of greatest importance to us as individuals and social beings are speech and music. Just like pure tones, speech and music are sound waves which impinge on the ear. But unlike pure tones, speech and music have come to convey to us an almost infinite variety of meanings and emotional experiences. It is not surprising, then, that the stimuli to such experiences are highly complex and continuously changing in time. In this section we shall be concerned with some of the relations which have proved of importance in communication and the perception of speech.

Speech as a Stimulus. Fig. 17 shows a picture of the acoustic waves as they emanate from a talker as he says, "Joe took father's shoe bench out." (This sentence is part of a standard utterance frequently used in the measurement of speech.) This pattern shows the wave-form of speech as a stimulus. Clearly, this stimulus rapidly varies both in frequency and intensity characteristics. We cannot

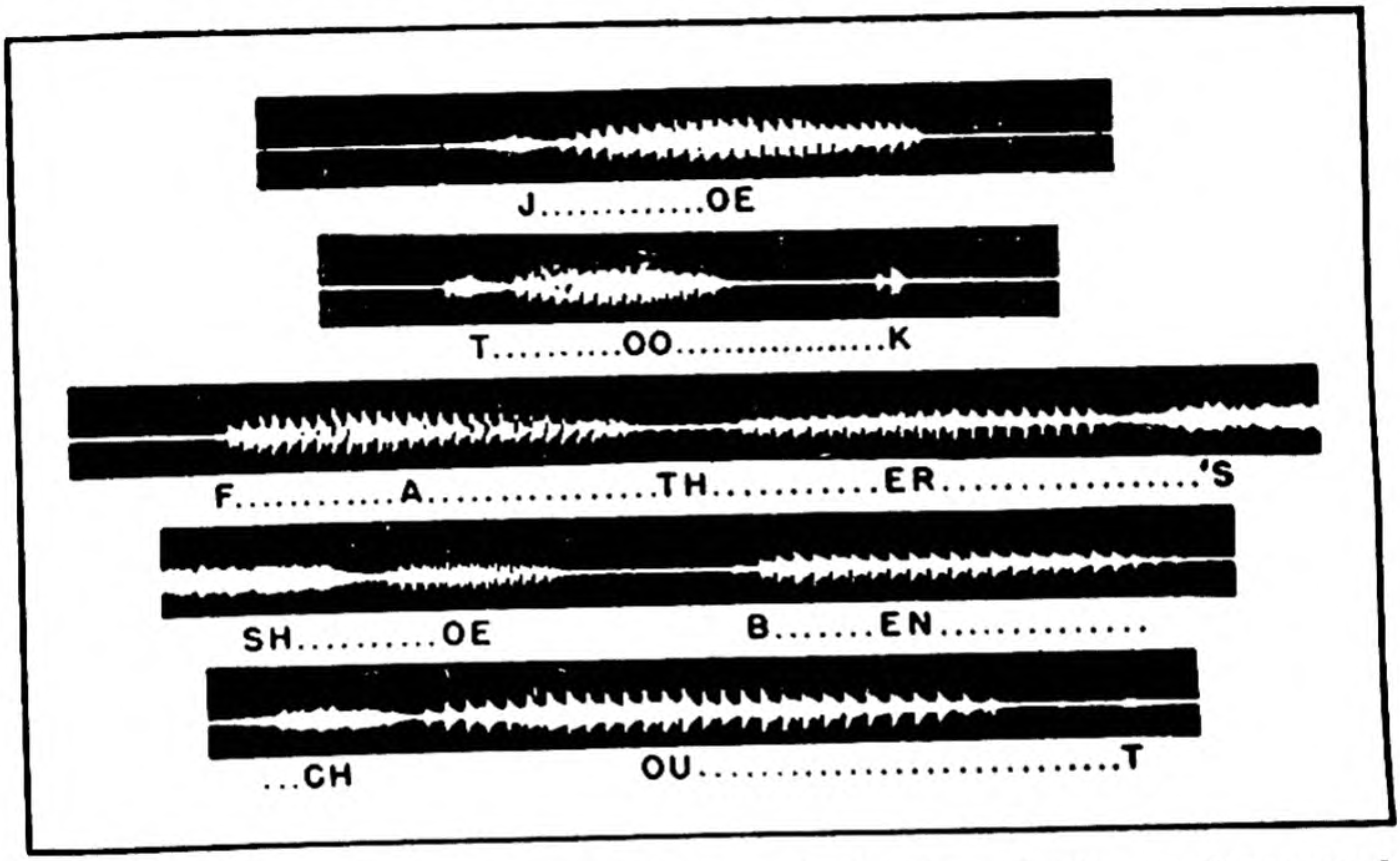


FIG. 17. An Oscillographic Picture of the Sound Waves Generated by a Speaker Pronouncing the Sentence, "Joe took father's shoe bench out." (From J. C. R. Licklider, D. Bindra, and I. Pollack. The intelligibility of rectangular speed waves, *Amer. J. Psychol.*, 1948, *61*:2, by permission of the journal and authors.)

tell by simple examination what it is about these speech waves that carries the "intelligibility" of a word. It is only by psychophysical analysis that we can relate characteristics of the speech stimulus to the perceived word.

Articulation Testing. First of all, a quantitative measure of the intelligibility of speech is required if we are to determine the factors influencing it. Such a quantitative measure may be obtained by simply counting the number of discrete speech units which a listener recognizes under specified conditions. A talker pronounces

selected speech units, and listeners record what they hear. Such a procedure is called an *articulation test.* Let us now illustrate some of the results which this procedure has provided.

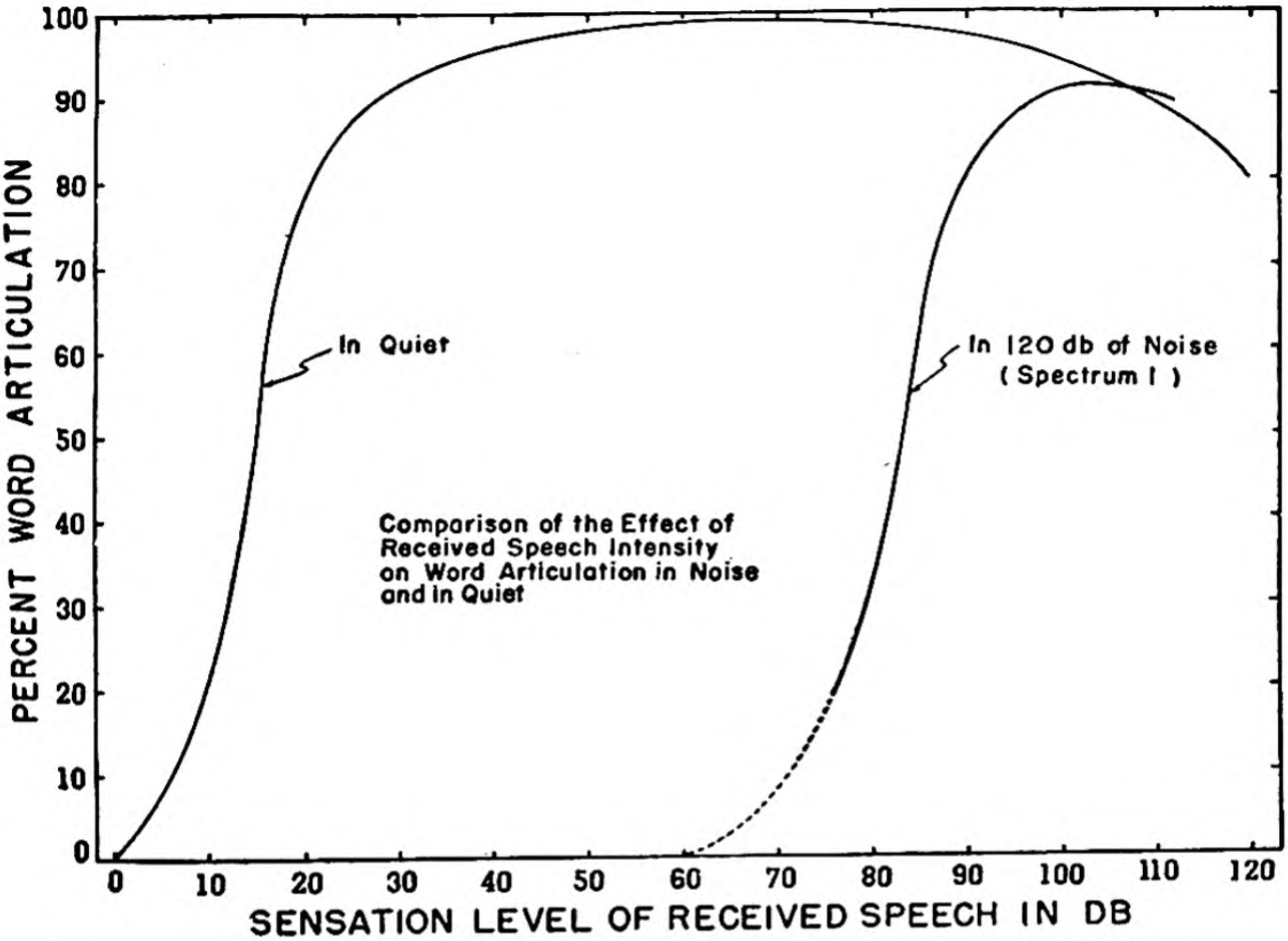


FIG. 18. Intelligibility of Speech in Quiet and in Noise. The curves show the percentage of words correctly understood as a function of the intensity of received speech. The left-hand function was obtained in the quiet; the right-hand function in 120 db of masking noise. The distance between the two curves indicates the increases in the intensity of speech required to overcome the effects of masking noise. (NDRC Research on Sound Control Report, Articulation Testing Methods II, 1 November 1944, OSRD Report No. 3802, courtesy of Psycho-Acoustic Laboratory, Harvard University.)

Frequency Distortion. By means of articulation tests, it has been shown that many frequency components can be filtered out of speech without seriously interfering with its intelligibility. The middle range of speech frequencies are the most important ones for intelligibility, and the lows and the highs may be missing without

rendering the words unintelligible. It is true that we occasionally cannot tell whether our telephone receiver has said "s" or "f" because there is no transduction of the high frequency components by this instrument. In these cases, however, context may supply the meaning, and make up for what is lost by frequency distortion.

Masking of Speech. The articulation test has also served as a useful procedure in the evaluation of the effectiveness of different noises in masking speech. We almost invariably listen to speech in the presence of noise. Most of the time these noises do not disturb us, and we usually have the time to repeat ourselves if we are not understood. Sometimes the noise is so intense that words are drowned out, and the perception of speech is impossible. In noisy vehicles, such as subways, airplanes, and tanks, communication is difficult indeed. If we are to learn how to overcome the intense masking noises, we must investigate the ways in which noise masks speech. Let us briefly illustrate. The weaker the intensity of a speech sound, the less likely it is to be understood correctly. Consider the left-hand curve of Fig. 18. This function shows that speech heard in the quiet and near the absolute threshold is quite unintelligible, but that as the intensity of the received speech is increased, more and more of the words are understood. Moreover, this curve shows that after a certain medium intensity has been reached, maximal intelligibility is obtained. At high intensities, there is a slight decline in the number of words understood.

The right-hand curve in Fig. 18 shows a similar function obtained with the talker and the listeners in an intense ambient noise. The intensity of received speech at which words begin to be intelligible must now be greater by 60 db than it was in the quiet. The threshold for hearing was greatly increased by the noise. Starting from the new base line, a function similar to that in the quiet is obtained. By making the speech sounds loud enough, the effects of the masking noise can be largely overcome.

Special Problems of Control in Auditory Experiments

How Quiet Must the Experimental Room Be? Not every auditory experiment requires absolute silence. When earphones are worn and especially when the test stimuli are well above the abso-

lute threshold, a certain amount of room noise can be tolerated. On the other hand, if the stimuli are weak or at the absolute threshold, a soundproofed room provides the ideal conditions. The question that must always be asked is: How much masking by extraneous stimuli can be tolerated without affecting the experimental results?

Interruption of Test Tone. When the intensity of a tone is decreased in order to determine the absolute threshold, it is important that the tone be frequently interrupted. If the tone sounds continuously as the intensity decreases, it becomes difficult for the subject to maintain a sharp distinction between presence and absence of tone. The need to interrupt the tone is not limited to measurements of the absolute threshold. In fact, in most auditory experiments, periodic interruptions of the tone help the subject to reestablish the criteria of judgment.

Control of Clicks. A pure tone cannot be turned on abruptly without producing an audible click. Suppose the oscillator is set at 1000 c.p.s. When the switch is thrown and the electrical wave is applied to the earphone, the diaphragm vibrates in a complex manner and thus produces a complex sound wave which the ear hears as a click. Thereafter, the vibrations of the diaphragm follow the frequency of the oscillator. Similarly, when the tone is turned off, a click is produced. These clicks may interfere with the subject's judgment. The harmful effect of clicks may be avoided by using a device which turns the tones on and off gradually rather than abruptly. If it takes about 0.2 seconds for a tone to reach its maximum intensity, little or no click will be heard.

Age of Subjects. In many psychological experiments in which adults are used as subjects, their ages are not, within wide limits, of crucial significance. In auditory experiments, however, it is important to pay careful attention to age in selecting subjects. As we grow older, our sensitivity to weak sounds decreases. This loss in sensitivity as a function of age is most marked for tones of high frequency. For example, at 8000 c.p.s., the threshold of a listener in his fifties is about 25 db higher than that of a twenty-year-old. It is for this reason that the ears of young subjects are used for the calibration of audiometers. Whenever the intensity of a tone is expressed in terms of sensation level, the subject's age becomes an important parameter.

Experiment III
The Threshold of Hearing

The sensitivity of the ear varies with the frequency of the stimulus tone as well as with the individual. Normal hearing consists of the ability to perceive a given tone at the sound-pressure level required by the so-called average subject. A loss in sensitivity usually means, therefore, that the subject requires a more intense stimulus than the average subject if he is to hear a given tone. Losses in sensitivity often involve only a limited number of frequencies rather than the entire audible range.

Fig. 19. An Audiometer in Use.

Purpose. To measure a subject's sensitivity to pure tones as compared with the norms established by measurements on normal, young adults. This experiment has great practical importance, since it is one of the standard procedures in the diagnosis of hearing defects.

Apparatus. An audiometer is desirable for this experiment. An audiometer is a device for generating pure tones of various frequencies and intensities. The sensitivity of each ear is measured separately by use of a single earphone connected to the audiometer. This instrument produces pure tones whose frequencies cover the important range of normal hearing. The intensity scale is so calibrated that at each frequency a setting

of zero corresponds to the absolute threshold of the normal ear. The positive numbers above zero indicate hearing loss in decibels, i.e., how many decibels above the average threshold a tone must be in order that the subject may hear it. The negative numbers similarly indicate better than average sensitivity. Fig. 19 shows a picture of an audiometer in use.

Procedure. It is essential that this experiment be performed in a quiet room. The subject is seated in such a position that he cannot see

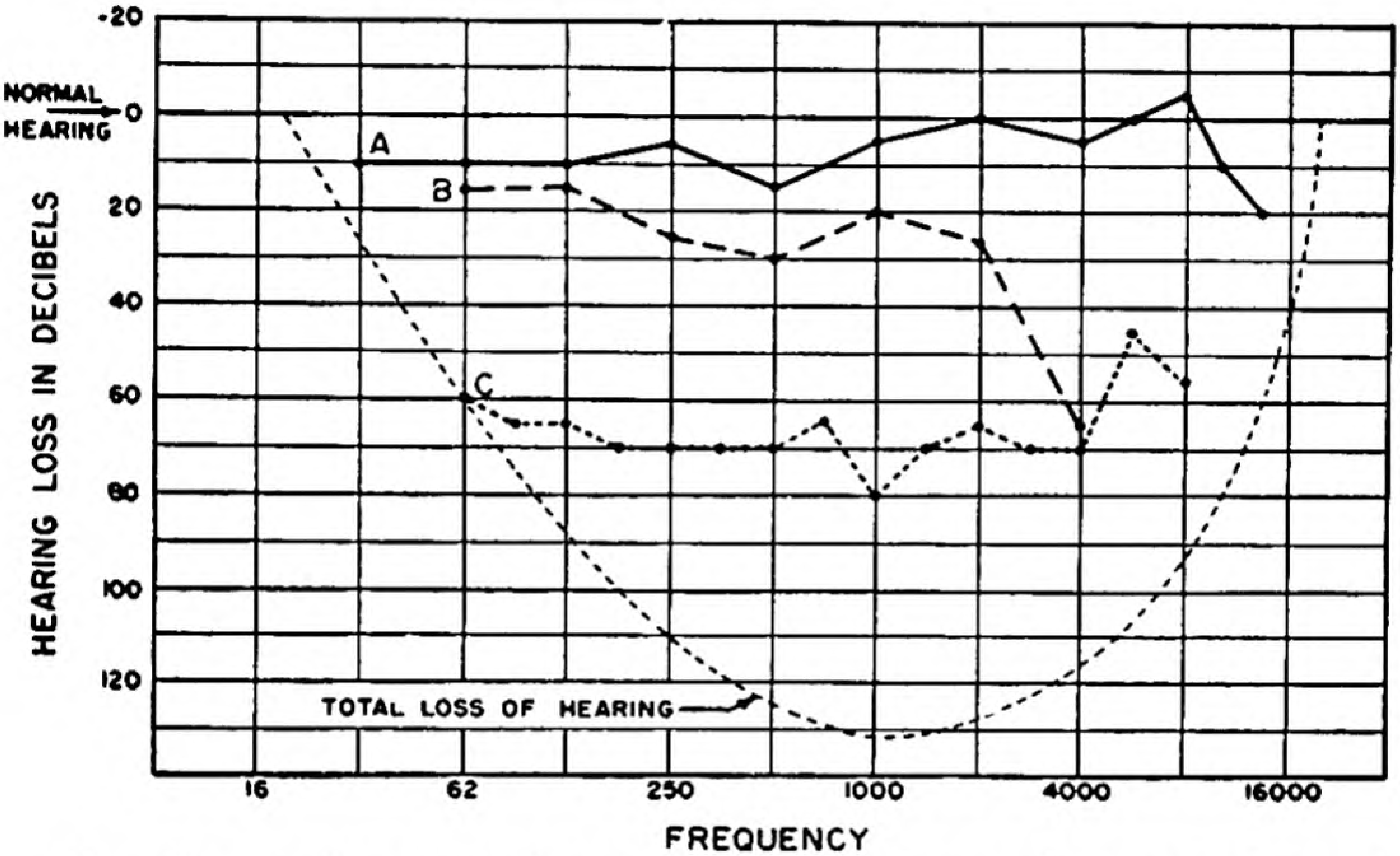


Fig. 20. Audiograms Showing Different Types of Hearing Loss. A: Hearing loss for low frequencies and normal hearing for frequencies above 1000 cycles. Note dip at very high frequencies. B: Same loss for low frequencies and increasing loss for higher frequencies. The dotted curve represents total loss of hearing. (From E. G. Boring (ed.), *Psychology for the armed services,* 1945, p. 108, by permission of Infantry Journal, Inc.)

the operation of the audiometer and he is instructed to hold the receiver snugly to his ear. He is provided with a push-button switch to signal whenever he hears a tone. The push-button switch operates a signal lamp on the face of the audiometer. The experimenter turns the frequency dial to a given frequency. It is desirable to begin with an intermediate frequency, say, 1000 c.p.s. The intensity of the tone is adjusted to a value which the subject can easily hear, and then is decreased until he fails to respond. The process is repeated through a smaller range—starting with a

smaller intensity and then decreasing it again until the subject no longer can hear the tone. Occasionally, the tone should be interrupted by means of a switch provided for that purpose. In this way we make sure that the subject signals only when he hears the tone. After several trials, the experimenter estimates as closely as possible the intensity above which the subject consistently responds and below which he fails to respond. The procedure is repeated for each of the frequencies desired. Usually about nine frequencies ranging from 100 to 8000 c.p.s. are adequate for the purposes of the experiment.

After the measurement of each threshold, the experimenter may enter the amount of hearing loss in decibels on an *audiometric chart*. The curve connecting the points on the chart is called an *audiogram*. An audiogram is reproduced in Fig. 20.

Frequently a subject may have normal hearing over most of the audible range but have a sharp *dip* over a restricted part of the audiogram as is shown in the figure. Such a restricted hearing loss usually occurs in the region of high frequencies.

Experiment IV
Masking

Whereas Experiment III is concerned with a subject's sensitivity in the quiet, this experiment shows the effect of a noise on the subject's absolute threshold. The results of such experiments have practical significance since, in many situations, communication takes place in noisy surroundings. For example, the navigator in an airplane must receive messages from the airport with the roar of the propeller in his ears. It is not that he is distracted by this noise, for he could overcome distraction by concentrated effort. The sounds that are masked by the propeller do not function as stimuli. As in the case of the navigator, most of the masking and masked stimuli in practical situations are highly complex. For purposes of experimental analysis, however, it is most useful to employ pure tones as the test stimuli. The general quantitative relationships established with pure tones have been successfully applied to situations in which complex stimuli, such as speech, are masked.

Purpose. To measure the shift in the absolute threshold due to the presence of a masking noise.

Apparatus. Two sources of sound are necessary: one, to generate pure tones; the other, to produce a masking noise. For the pure tones, an audio-oscillator is desirable. Also needed is a device for controlling the intensity of the output of the oscillator. An attenuator calibrated to read in decibels is most convenient. Either earphones or a loudspeaker may be

used to present the test tone. If an audio-oscillator is not available, an audiometer may be substituted. The masking noise may be produced in various ways. An electrical source of noise is best since its intensity is readily controlled. However, any other device, such as a loud buzzer or a motor with gear attachments, will do.

Procedure. The experiment is begun in a quiet room. The subject's absolute threshold is determined at a given frequency, say, 1000 c.p.s. Then the noise stimulus is turned on, and the subject's threshold for the pure tone is again determined (masked threshold). The difference between the masked threshold and the absolute threshold expressed in decibels is the amount of masking. This basic experimental procedure may be extended in two ways. If the experimenter can control the intensity of the masking noise, then he can determine the amount of masking as a function of the sound-pressure level or sensation level of the noise. Moreover, it is possible to change the frequency of the test tone and to determine masking as a function of frequency with a constant level of the masking noise.

If the masking noise is not delivered directly to the earphones, a special problem of interpretation arises. To a certain extent, the earphones insulate the subject's ear from a noise produced in the room. Under these conditions, less masking will be found than if the noise fell directly on the unprotected ear. Of course, if the test tone is produced by a loudspeaker, this consideration does not arise. Presentation of the test tone by loudspeaker, however, has a disadvantage. If the walls of the room have not been sound-treated, even slight movements of the head may affect the obtained threshold values.

General References

Boring, E. G., *Sensation and perception in the history of experimental psychology,* New York: Appleton-Century and Co., 1942, Chaps. 9, 10, and 11.

Fletcher, H., *Speech and hearing,* New York: D. Van Nostrand Co., 1929.

Morgan, C. T., *Physiological psychology,* New York: McGraw-Hill Book Co., 1943, Chap. 11.

Stevens, S. S., and Davis, H., *Hearings: its psychology and physiology,* New York: John Wiley and Sons, 1938.

Wever, E. G., Studying hearing. In T. G. Andrews (ed.), *Methods of psychology,* New York: John Wiley and Sons, 1948.

Auditory Stimulus

Miller, D. C., *The science of musical sounds,* New York: The Macmillan Co., 1922.

Stevens, S. S., and Gerbrands, R., A twin-oscillator for auditory research, *Amer. J. Psychol.*, 1937, *49*:113-115.

Wood, A. B., *Acoustics,* London: Blackie and Son, 1940.

Auditory Discrimination

Knudsen, V. O., The sensibility of the ear to small differences in intensity and frequency, *Phys. Rev.*, 1923, *21*:84-102.

Riesz, R. R., Differential intensity sensitivity of the ear for pure tones, *Phys. Rev.*, 1928, *31*:867-875.

Shower, E. G., and Biddulph, R., Differential pitch sensitivity of the ear, *J. Acous. Soc. Amer.*, 1931, *3*:275-287.

Sivian, L. J., and White, S. D., On minimum audible sound fields, *J. Acous. Soc. Amer.*, 1933, *4*:288-321.

Attributes of Auditory Experience

Fletcher, H., and Munson, W. A., Loudness, its definition, measurement, and calculation, *J. Acous. Soc. Amer.*, 1933, *5*:82-108.

Stevens, S. S., The attributes of tones, *Proc. Nat. Acad. Sci.*, 1934, *20*:457-459.

Stevens, S. S., The volume and intensity of tones, *Amer. J. Psychol.*, 1934, *46*:397-408.

Stevens, S. S., The relation of pitch to intensity, *J. Acous. Soc. Amer.*, 1935, *6*:150-154.

Stevens, S. S., A scale for the measurement of a psychological magnitude: loudness, *Psychol. Rev.*, 1936, *43*:405-416.

Stevens, S. S., and Davis, H., Psychophysiological acoustics: pitch and loudness, *J. Acous. Soc. Amer.*, 1936, *8*:1-13.

Stevens, S. S., and Volkmann, J., The relation of pitch to frequency: a revised scale, *Amer. J. Psychol.*, 1940, *53*:329-353.

Physiological Basis of Pitch and Loudness

Culler, E. A., In Symposium on tone localization in the cochlea, *Ann. Otol., Rhinol. and Laryngol.*, 1935, *44*:809-815.

Davis, H., Derbyshire, A. J., Kemp., E. H., Lurie, M. H., and Upton, M., Functional and histological changes in the cochlea of the guinea-pig resulting from prolonged stimulation, *J. Gen. Psychol.*, 1935, *12*:251-278.

Galambos, R., and Davis, H., The response of single auditory-nerve fibers to acoustic stimulation, *J. Neurophysiol.*, 1943, *6*:39-57.

Stevens, S. S., and Davis, H., Psychophysiological acoustics: pitch and loudness, *J. Acous. Soc. Amer.*, 1936, *8*:1-13.

Stevens, S. S., Davis, H., and Lurie, M. H., The localization of pitch perception on the basilar membrane, *J. Gen. Psychol.*, 1935, *13*:297-315.

Wever, E. G., The physiology of hearing: the nature of the response in the cochlea, *Physiol. Rev.*, 1933, *13*:400-425.

Woolsey, C. N., and Walzl, E. M., Topical projection of nerve fibers from local regions of the cochlea to the cerebral cortex of the cat, *Johns Hopkins Hosp. Bull.*, 1942, *71*:315-344.

Beats, Difference Tones, Masking, and Aural Harmonics

Fletcher, H., Loudness, pitch and the timbre of musical tones and their relation to the intensity, the frequency and the overtone structure, *J. Acous. Soc. Amer.*, 1934, *6*:59-69.

Stevens, S. S., and Newman, E. B., On the nature of aural harmonics, *Proc. Nat. Acad. Sci.*, 1936, *22*:668-672.

Wegel, R. L., and Lane, C. E., The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear, *Phys. Rev.*, 1924, *23*:266-285.

Wever, E. G., Beats and related phenomena resulting from the simultaneous sounding of two tones, *Psychol. Rev.*, 1929, *36*:402-418; 512-523.

Wever, E. G., The designation of combination tones, *Psychol. Rev.*, 1941, *48*:93-104.

Auditory Localization

Steinberg, J. C., and Snow, W. B., Physical factors in auditory perspective, *Bell System Tech. J.*, 1934, *13*:245-258.

Stevens, S. S., and Newman, E. B., The localization of actual sources of sound, *Amer. J. Psychol.*, 1936, *48*:297-306.

Wallach, H., The role of head movements and vestibular and visual cues in sound localization, *J. Exper. Psychol.*, 1940, *27*:339-368.

Wightman, E. R., and Firestone, F. A., Binaural localization of pure tones, *J. Acous. Soc. Amer.*, 1930, *2*:271-280.

Auditory Fatigue

Banister, H., Audition: I. Auditory phenomena and their stimulus corre-

lates. In C. Murchison (ed.), *A handbook of general experimental psychology*, Worcester: Clark University Press, 1934.

Rawdon-Smith, A. F., Auditory fatigue, *Brit. J. Psychol.*, 1934, 25:77-85.

Rawdon-Smith, A. F., Experimental deafness. Further data upon the phenomenon of so-called auditory fatigue, *Brit. J. Psychol.*, 1936, *26*:233-244.

SPEECH, HEARING, AND COMMUNICATION

Egan, J. P., Articulation testing methods, *The Laryngoscope*, 1948, *58*: 955-991.

Fletcher, H., *Speech and Hearing*, New York: D. Van Nostrand Co., 1929.

NDRC Research on Sound Control report, *Articulation testing methods: II*, 1 November 1944, OSRD, Report No. 3802.

»»» 5 «««

SMELL AND TASTE

TWO of our senses may well be treated together: smell and taste both respond to chemical changes, and the specialized receptors of these two sensory systems are sometimes described as chemoreceptors. The common action of smell and taste is clearly brought out when we analyze the ways in which we perceive the flavor of food. Although we speak of tasty or tasteless dishes, most of the flavor comes from olfactory stimulation and not from the relatively impoverished sense of taste.

In this chapter we shall briefly summarize the principal psychophysical facts pertaining to these two senses. It should be said at the outset that the experimental psychology of smell and taste lags far behind the exploration of the other major senses. The great advances made in the psychophysics of the other major senses stem from our ability to specify and control the dimensions of the stimulus with which the attributes of experience are correlated. In the case of smell and taste, the physical dimensions effective in stimulation have not been isolated. It is true that we know what the objects are that are tasted and smelled, but we cannot usually state what particular characteristics of the stimulus object systematically determine variations in olfactory and gustatory experience. For this reason there is less to say about smell and taste than about the other senses, and stable quantitative generalizations are rare.

SMELL

The Olfactory Stimulus

Nature of the Olfactory Stimulus. • Man sees when light waves strike his retina. He hears when sound waves are transmitted to the inner ear. As we have pointed out above, we cannot specify a par-

ticular class of stimulus events which are a necessary condition for smelling. We do know, however, that only contact with gaseous particles provides the adequate stimulation for the specialized smell receptors. These smell receptors are located in the olfactory membrane in the upper cavity of the nose. When we inhale, most of the air does not reach the olfactory membrane, and it is principally by eddy currents that the gaseous particles necessary for stimulation are carried to the receptors. Sniffing aids in this process. The fact that it is gaseous particles and not solutions which are the adequate stimuli for smell was first indicated in a classical experiment performed a long time ago. The investigator filled his nasal cavity with

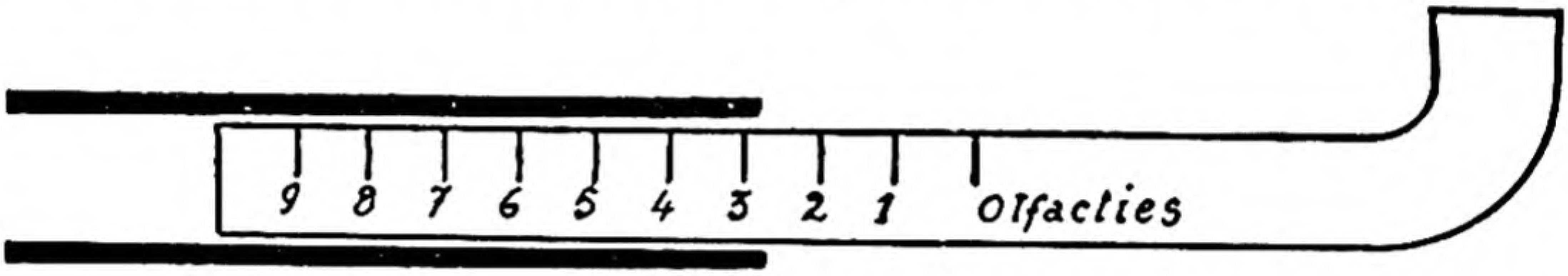


FIG. 21. An Olfactometer Calibrated in Olfacties. The odorous substance is placed on the inner surface of the outer tube. The farther out the inner tube is pulled, the larger the amount of odorous substance exposed to air and hence the more intense the olfactory stimulus. (From W. J. Crozier, Chemoreception. In C. Murchison (ed.), *A handbook of general experimental psychology*, 1934, p. 999, by permission of Clark University Press. After Zwaardemaker, 1895.)

eau de Cologne. He reported that under these conditions he could not smell this substance notorious for its odor. This conclusion was confirmed by later experimenters.

Techniques of Stimulation. One of the difficulties with which investigators of olfaction have had to contend is the inaccessible location of the sense organ. It is very hard, indeed, to apply a known quantity of odorous substance to the receptor cells. Many methods of stimulation have been devised. We shall briefly discuss some of these.

The best known stimulating device is the *olfactometer,* illustrated in Fig. 21. This instrument consists of two tubes, constructed so that one may slide over the other. The smaller of the two tubes terminates in a nosepiece which is inserted into the nostril. On the inner surface of the larger tube is placed the odorous substance, for example, india rubber. By sliding out the smaller tube, the

stream of air is allowed to pass over a portion of the odorous substance. The farther out the smaller tube is pulled, the larger the portion of substance exposed to the current of air and, hence, the number of particles which presumably reach the sense organ. When the outer tube covers the inner tube, there is presumably no stimulation. The distance by which the inner tube must be withdrawn before the presence of an odor is detected is defined as an *olfactie.* This distance, defining an olfactie, is then taken as the unit for measurement of the intensity of the odorous stimulus. When the inner tube is withdrawn twice the threshold distance, we say that we have two olfacties. Thus, this method of stimulation allows a semi-quantitative specification of the stimulus. The method is semi-quantitative because it is not possible to control accurately the amount of air passing over the odorous substance every time the subject takes a sniff. Nevertheless, the olfactometer has been a useful instrument. It has been possible to adapt it to liquid stimulus substances by mixing the liquid with melted paraffin and then forming a rod from the mixture. In a variety of experimental situations a double olfactometer is employed, one for each nostril. In this way different amounts of the same odorous substance as well as different substances can be introduced separately into the nostrils.

A method which allows greater control over the intensity of the stimulus is the *blast injection test.* With this method, an odorous gas under a given, constant pressure can be injected into the nasal cavity at a uniform rate. The subject holds his breath while the injection is made. When this method is employed, the smallest amount of air needed for the detection of an odor defines the absolute threshold. Of course, the air pressure and the concentration of the odorous substance must be specified.

The stimulus can be further controlled by making determinations in an odorless environment. A laboratory room has many sources of odor which are difficult to control. For this reason, the *camera inodorata* was invented. This device merely consists of an enclosure with glass sides and aluminum top and bottom, with an adjustable opening through which the subject puts his head. The air in this box is deodorized with the aid of a mercury-vapor lamp. Determinations in the camera inodorata result in considerably lower thresholds than those made in the air of the room.

Several other techniques of stimulation have been employed, such as sniffing from an open bottle, leading a tube from a stoppered bottle directly to the nose, etc. It is certainly true that control of stimulus intensity is neither standardized nor technically perfected.

Another problem of stimulus control should be mentioned here. The threshold values may be influenced by the temperature of the substance used in stimulation. Since the threshold may vary appreciably with the temperature, this factor should always be specified and taken into account in comparing experimental findings.

Psychophysics of Olfaction

Absolute Thresholds. Using one of the techniques of stimulation described above, the absolute thresholds for a large variety of odorous substances have been determined. It turns out that the olfactory system is an extremely sensitive one and will detect astonishingly small concentrations of odorous substances. For example, the odor of vanillin, the fragrant constituent of vanilla, can be detected when its concentration is only 0.0000002 milligrams per cubic meter of air. Such a specification of the threshold is in terms of weight of the odorous substance per volume of air. Another way to specify the absolute threshold is in terms of number of molecules of the odorous substance per cubic centimeter at threshold. Here are a few representative values:

ethyl alcohol	3×10^{13}
benzene	4×10^{11}
cresol	5×10^{10}
iodoform	4×10^{8}

These numbers, although large, still are exceedingly small in the frame of reference of molecular numbers. It must, moreover, be remembered that many of the molecules presented never reach the receptors.

Differential Thresholds. Differential sensitivity to the intensity of odor can be determined by one of the usual psychophysical methods. The subject is required to discriminate which of two odors is the more intense. Subjects can make such discriminations, but differential sensitivity is quite poor as compared with some of the other sensory systems, such as proprioception and vision. An increase

in concentration of about 30 percent may be required before the subject can tell a difference. The differential sensitivity varies with the intensity level at which the determinations are made. As in other

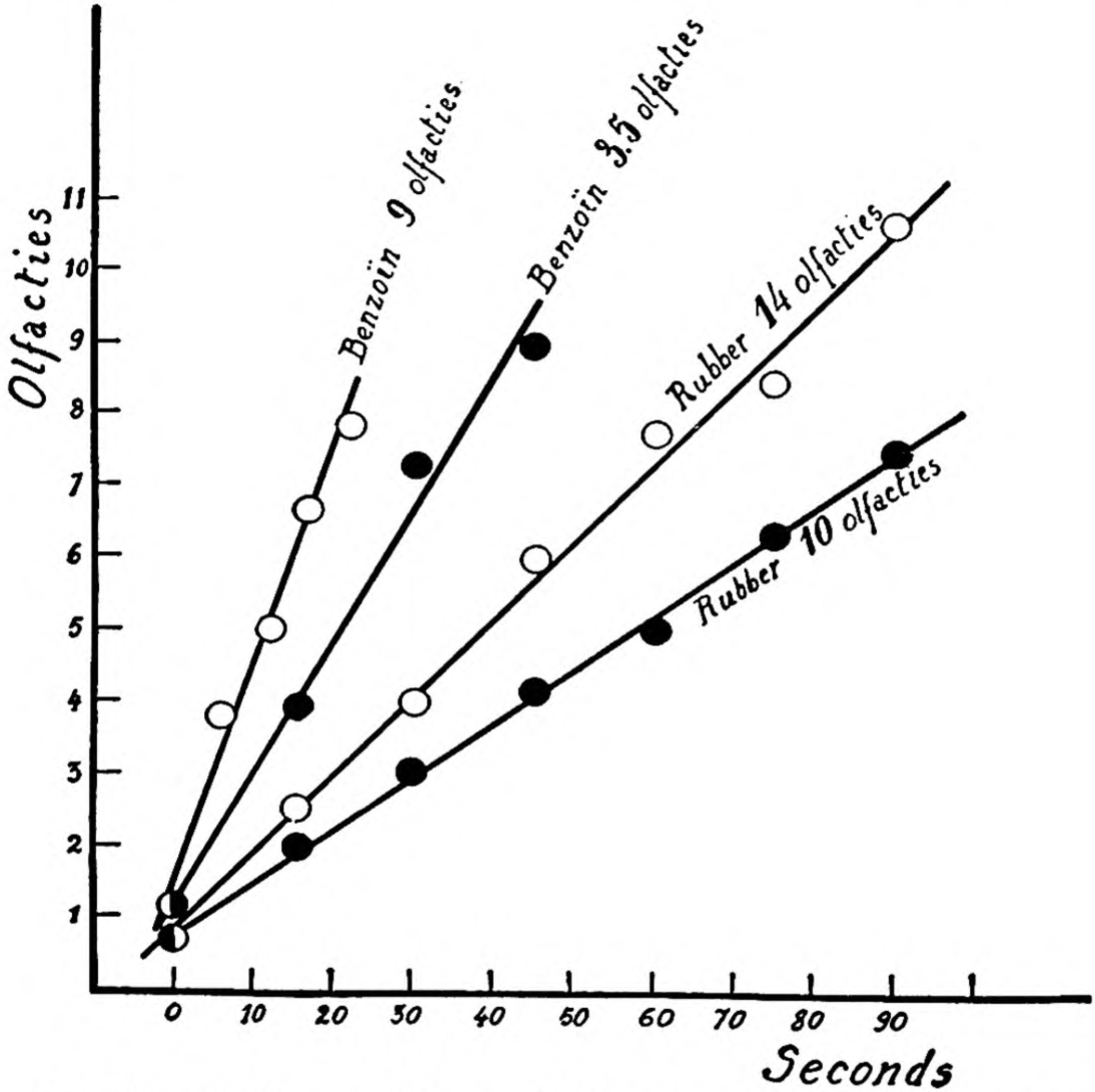


FIG. 22. The Temporal Course of Adaptation to an Olfactory Stimulus as Determined by the Specific Nature and the Intensity of the Stimulus. Each point represents determinations of the absolute threshold after varying lengths of exposure to the stimulus. (From W. J. Crozier, Chemoreception. In C. Murchison (ed.), *A handbook of general experimental psychology,* 1934, p. 1002, by permission of Clark University Press. After Zwaardemaker, 1895.)

sense modalities, relative sensitivity to change increases as intensity is increased.

Adaptation. Sensitivity to odors also changes as a function of the length of exposure to a stimulus. As in other sense modalities, continued stimulation leads to a decrease in sensitivity, i.e., there is adaptation. Olfactory adaptation may be complete in a matter of minutes. For example, after 3 minutes of exposure to oil of lemon, the perception of the odor may have completely disappeared. If we are interested in tracing the course of adaptation in quantitative terms, we may do so by measuring the absolute threshold after varying lengths of exposure to the stimulus. Fig. 22 shows the results of such an investigation. These data indicate that the rate of adaptation is jointly determined by the specific nature of the odorous substance and its intensity.

Adaptation is not restricted to the specific odor of the adapting stimulus, but may increase the threshold for a number of other odorous stimuli. For example, adaptation to ammonium sulphide raises the threshold for that substance and for hydrogen sulphide, but not for ethereal oils. It was hoped that such *selectivity* of adaptation might provide an avenue to the classification of odorous substances. Unfortunately, this hope has not been realized.

In the course of adaptation, there may be not only a decrease in sensitivity but also a qualitative change in the experience. This *modulation* of odor is usually illustrated by pointing out that cheap perfumes smell agreeable at first, but gradually grow more and more disagreeable as stimulation continues.

Qualities of Odor. It is easier to measure changes in sensitivity to odor than to arrive at a classification of odor qualities. The variety of odors to which man is sensitive is immense. There are so many blends and nuances that it almost seems as though there are as many specific odors as there are substances that smell. Some of these odors resemble each other more than they do others. It is natural, then, that there has been a concerted effort to classify odors and reduce the manifold of experience of odor to a few primary categories from which the blends and nuances could be compounded. The psychology of odor has aimed at the same goal that the psychology of color has so successfully achieved. For there is a rich diversity of color experiences, probably no less variegated than

the multiplicity of odors, and yet it is possible to specify a given color in terms of a few primary points of reference. The main attempts to find similar points of reference for odors have not been too successful. The one scheme usually cited as the most acceptable is the *smell prism,* shown in Fig. 23. This system of classification assumes that there are six primary odors, and that any odor may be classified as belonging to one of these six categories or as intermediate between them. Careful examination of Fig. 23 shows that the primaries are located at the corners of the prism:

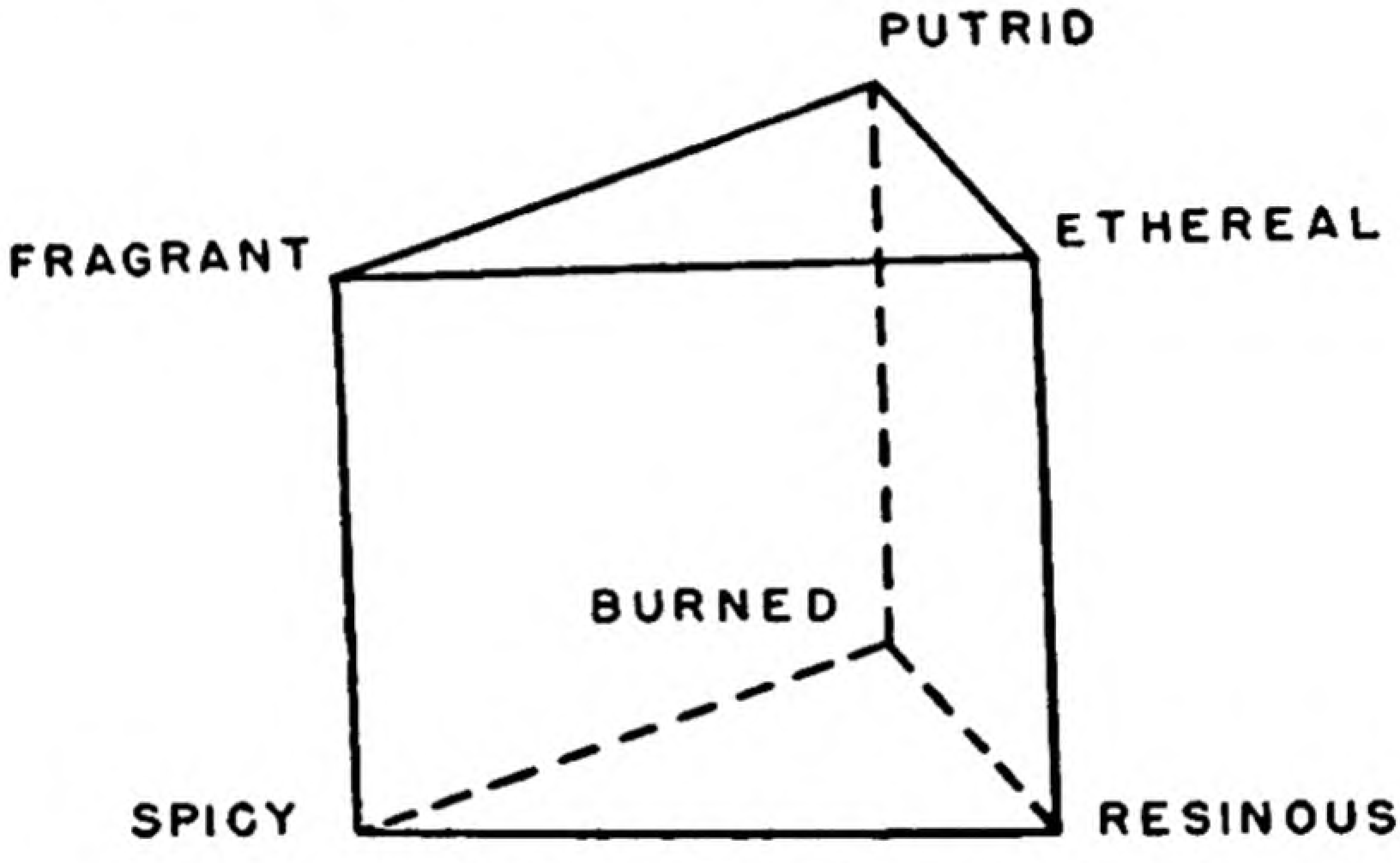


FIG. 23. A Scheme for the Classification of Odors: Henning's Smell Prism.

1. *Spicy,* cloves and cinnamon are examples.
2. *Fragrant,* the smell of many flowers, such as violets and geranium.
3. *Ethereal,* the odor of many fruits, such as oranges and lemons.
4. *Resinous,* the odor of balsam and turpentine.
5. *Burned,* of which the odor of tar is a good example.
6. *Putrid,* the smell of rot and decay of organic matter, e.g., H_2S.

All odors which do not fall into the primary categories are thought to be specific shadings and blends falling somewhere between them. They are localized along the edges and the outside surfaces of the smell prism. It was hoped it would be possible to proceed from

corner to corner in an orderly series. Examples of graded series of odors localized along the edges of one of the surfaces of the prism are shown in Fig. 24. It has not always been possible to obtain

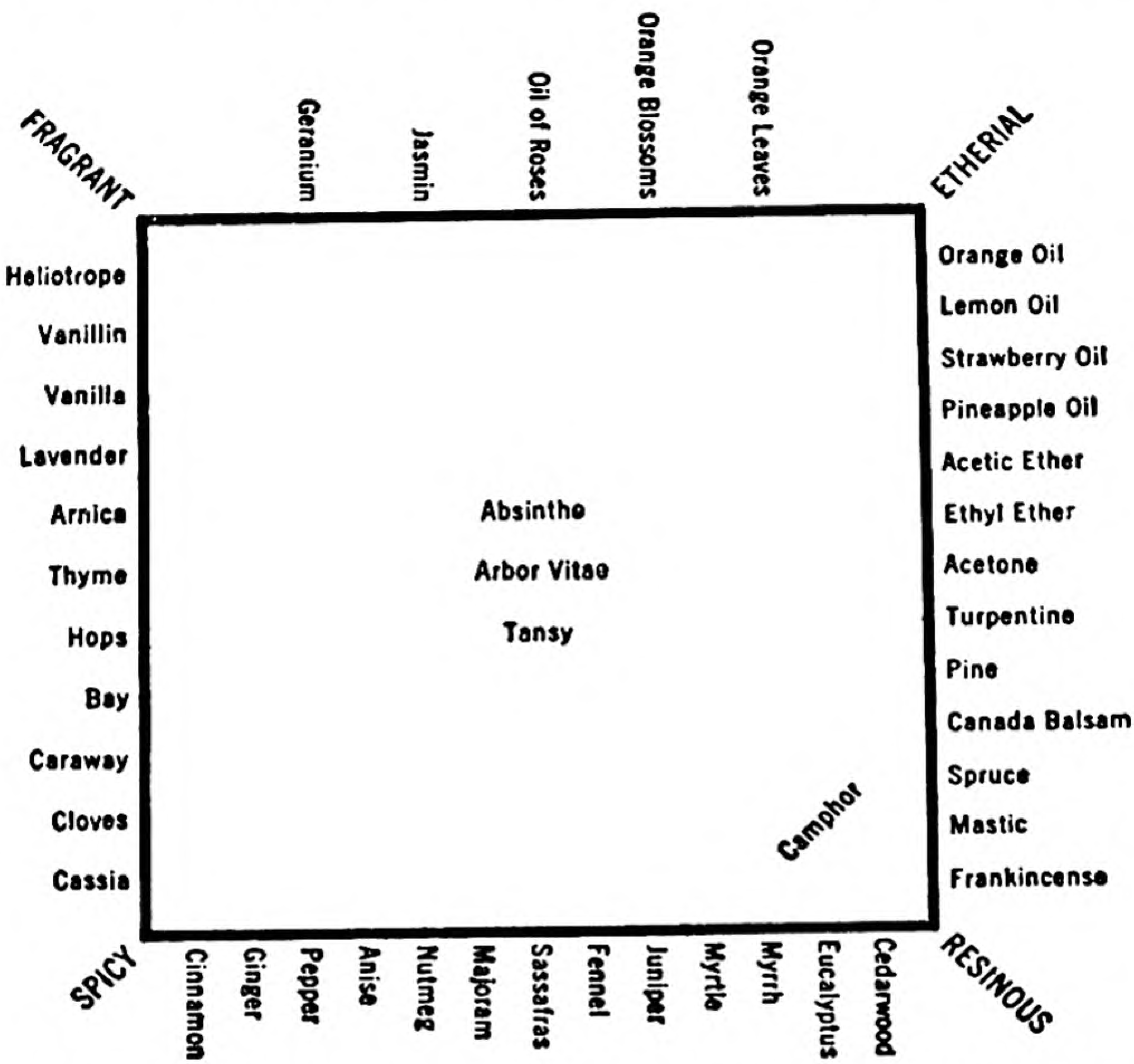


FIG. 24. A Graded Series of Odors Localized Along the Edges of One of the Surfaces of the Smell Prism Shown in Fig. 23. (From R. S. Woodworth, *Experimental psychology*, 1938, p. 486, by permission of Henry Holt and Company, Inc.)

agreement among different subjects as to the location of a specific odor on the prism, nor does the same subject always place the same odor in the same place. Nevertheless, this scheme of classification has its uses in bringing some order into the bewildering multiplicity of odor experiences, even though it is not a very reliable instrument for an accurate inventory of odors.

A great deal of research has been devoted to the chemical properties of olfactory stimuli as a potential avenue to a lawful classification of odors as well as to an understanding of the mode of action of the stimulus. This work has been largely concerned with organic compounds which constitute the majority of odorous substances. Now and again, promising relationships have emerged, such as the fact that members of a homologous series, e.g., the series of alcohols, tend to have similar qualities of odor. In general, however, there have been too many exceptions to permit generalizations about the correlation between the chemical properties of the stimulus and the qualities of odor.

Simultaneous Stimulation by Various Odors. When the sense organ is simultaneously stimulated by a number of odorous substances, several effects may take place. First of all, there are the *fusions,* to which we have already referred. When two substances are sniffed together, an olfactory blend may be perceived. By careful attention it may be possible to identify the presence of the two components. The more similar in odor the two components are, the more difficult it is to make this analysis. The quality of the fusion may change with adaptation. If one of the substances shows more rapid adaptation than the other, the identification of the more slowly adapting stimulus is made easier.

It sometimes happens that two odors in a mixture weaken each other so that the resulting experience is less intense than in the presence of either substance alone. This phenomenon is called *compensation,* and may be obtained most readily when both stimuli are weak. Some investigators have claimed that with carefully selected stimuli, presented at appropriate intensities, complete compensation or neutralization may occur, the two odors canceling each other out, as it were. Other experimenters have failed to confirm the occurrence of complete neutralization, and, in any event, this phenomenon is rare. How different this situation is from color mixture where every hue has its complementary!

Finally, there may be complete *masking* of one odor by another. Practically any odor may mask another odor provided the intensity difference between them is large enough. The principle of masking is extensively used to eliminate the unpleasant odors (the taste) of many medicines. Sometimes this takes a lot of doing as the fol-

lowing instance shows. To mask the notoriously repulsive odor of castor oil, it may be necessary to add as many as four other substances: saccharine, vanillin, oil of lemon, and oil of peppermint. With enough masking stimuli, even castor oil may become palatable. These phenomena are parallel to auditory masking. In auditory, as well as olfactory masking, one stimulus may prevent completely the perception of another (pp. 69-70).

Olfactory Localization. When only one of the nostrils is stimulated by an olfactory substance, the blindfolded subject cannot tell which one of his nostrils has been stimulated unless he is aided by other sensory information. Thus, the subject may localize the olfactory stimulus on the basis of a combination of smell with taste, warmth or cold, or pain. These results underline the close collaboration of smell with other sensory systems whose specialized receptors are located in the nose, throat, and mouth.

Taste

The Gustatory Stimulus

Nature of the Gustatory Stimulus. Taste is another chemical sense. The adequate stimulus to taste consists of certain substances in solution placed on the tongue. The tongue is equipped with taste cells grouped into taste buds, and these are the receptors that receive the gustatory stimulus. Typically, the effective stimuli for taste are substances soluble in water (or saliva), whereas in olfaction the substances are usually lipoid soluble. As in the case of smell, the precise chemical nature of the interaction between the sapid substance and the taste cells is not fully understood.

Techniques of Stimulation. Although more accessible than the olfactory cells, the gustatory receptors are embedded in the tongue, and a precise quantitative control of the gustatory stimulus still presents considerable difficulty. The specific technique of stimulation used varies with the size of the area that we desire to stimulate. For some purposes it is sufficient to ask the subject to sip a small amount of the solution, letting the liquid flow where it may on the tongue. If a series of such judgments is required, it is important to rinse the mouth carefully after each stimulation because "taste lingers" and adaptation may be rapid.

For many other purposes it is necessary to confine the stimulus to a narrowly circumscribed area of the tongue. A small brush or piece of cloth soaked in the stimulus substance is then employed. Recently, a mechanical device has been used which forces the liquid substance through small tubes onto the surface of the tongue and then quickly removes them. The same device may be used for rinsing the stimulated area.

Finally, important evidence about the nature of taste receptors has been obtained by applying a weak electric current to the tongue.

Whatever technique is used, the factor of temperature must be controlled. The sensitivity of the taste receptors varies in great measure with the temperature of the stimulus solution. Sweetened tea tastes sweeter as it cools. The temperatures for greatest sensitivity vary between 20° and 40° C.

Psychophysics of Taste

Absolute Thresholds. The absolute threshold is specified in terms of the concentration of the sapid substance that can just be detected. The accompanying list shows representative values for the commonly recognized qualities of taste. It will be noted that

Substance	Quality	Threshold Concentration (percent)
sugar	sweet	0.5
table salt	salty	0.25
hydrochloric acid	sour	0.007
quinine	bitter	0.00005

these values are listed from least to greatest sensitivity. We must not take this order as invariant, because it may be, and already has been, upset by tests of new substances. Thus, if saccharine is included in the above list, the order of sensitivity would be changed since a solution of only 0.0001 percent saccharine is detected as sweet.

The particular value for the absolute threshold for a given substance varies with the location on the tongue stimulated. Different areas of the tongue vary in their sensitivity to the four main types of gustatory stimuli. Greatest sensitivity to bitter is found at the back of the tongue, whereas sweet is most readily detected at the

tip. The lowest absolute thresholds to sour are obtained on the sides of the tongue, whereas the lowest thresholds for saline are distributed over the tip and sides. Some groups of cells (papillae) have been isolated which give rise to only one taste quality. In other cases, two or more of the primary taste qualities may result from stimulation by the appropriate substances.

The specific values of the absolute threshold also depend upon the duration of stimulus application. It is known that it takes a considerable amount of time for the stimulus to penetrate to the sensitive receptor cells. Even a supraliminal stimulus may take as long as 10 seconds to become fully effective.

The physiological state of the organism may become an important factor in the absolute sensitivity to sapid substances. It is well established that organisms whose adrenal glands are diseased or removed suffer a serious salt deficiency. This increase in need for salt is accompanied by a marked increase in the absolute sensitivity to salty stimuli. This increased sensitivity undoubtedly helps the animal to select food substances necessary for its survival.

Differential Thresholds. There are several reasons why differential sensitivity to the intensity of gustatory stimuli is hard to measure. If the experimenter wishes to present successively two stimuli for comparison, it is necessary to rinse the tongue, or the part stimulated, carefully between the application of the first and second stimulus. This necessary procedure makes it difficult to control the time interval between the two stimuli, and also makes the subject's task more difficult. The experimenter has to contend with the adaptation to the stimuli which may ensue while the determination is being made. It is not surprising, then, that the literature on this problem is not extensive. The work with saline solutions indicates, however, that the differential threshold with this stimulus is of the order of 25 percent for the solution yielding maximum sensitivity. Furthermore, the threshold values decrease as the concentration of salt becomes greater.

Adaptation. Continued application of a gustatory stimulus decreases the sensitivity to that stimulus, i.e., adaptation takes place. The temporal course of adaptation may be traced by stimulating the tongue for various periods of time with the adapting stimulus and determining the absolute threshold after each period. Fig. 25

shows that the threshold rises rapidly at first and, then, gradually. Also note that the greater the concentration of the adapting stimulus, the greater the rise in the threshold. This set of data also shows that after the removal of the adapting stimulus, sensitivity at first increases rapidly and then more slowly.

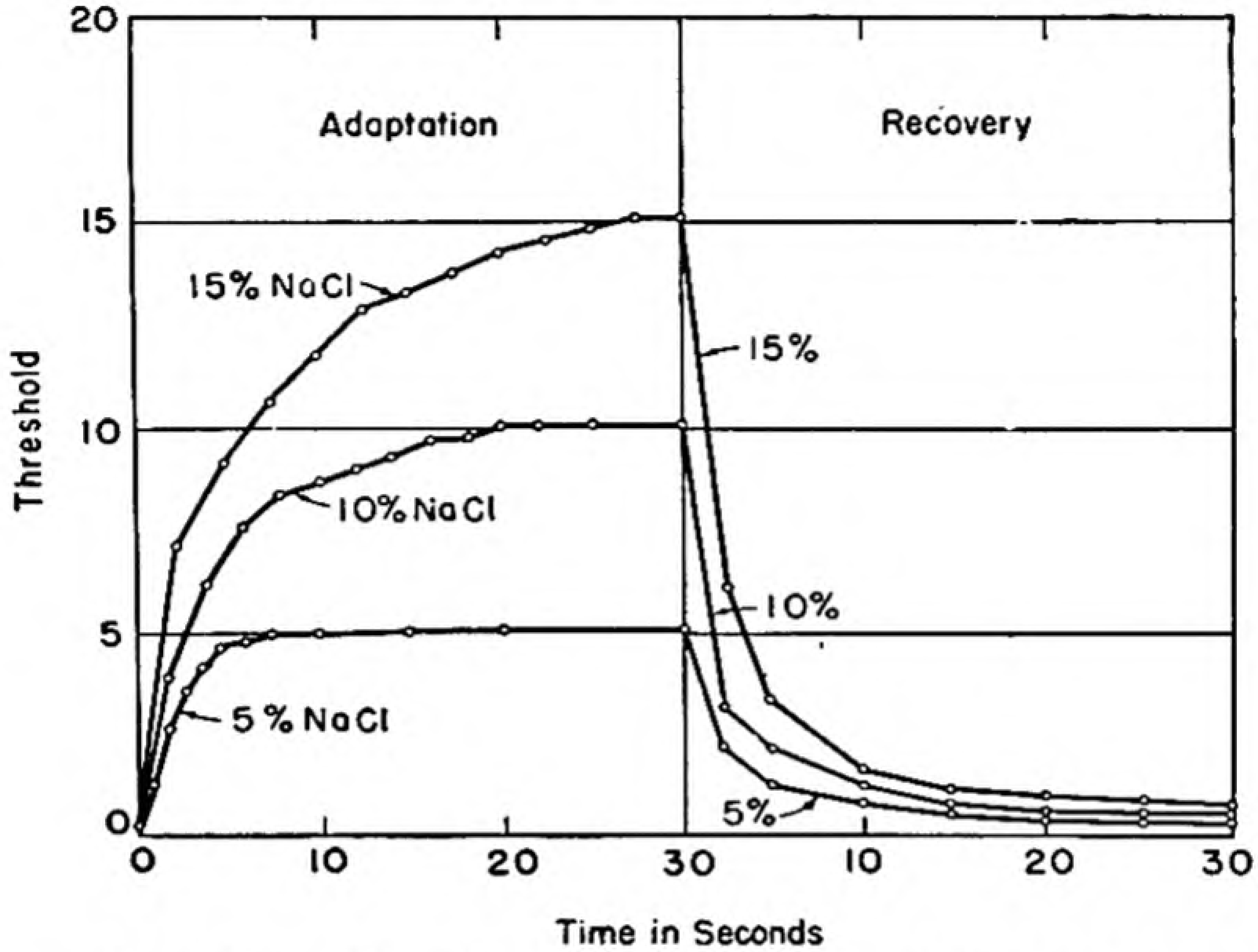


FIG. 25. The Temporal Course of Adaptation to Varying Intensities of a Gustatory Stimulus (salt) and the Course of Recovery of Sensitivity Following Removal of the Stimulus. Each point represents absolute thresholds determined at the time interval indicated on the abscissa. (Reproduced by permission from *Foundations of psychology* by Boring, Langfeld, and Weld, p. 354, published by John Wiley & Sons, Inc., 1948. After Hahn, 1934.)

An interesting question may be asked about adaptation to gustatory stimuli. If we adapt to one type of stimulus, say, a sweet one, does the absolute sensitivity to another type of stimulus, say, a salty one, undergo any change? One answer to this question has been that adaptation to a sweet stimulus increases the absolute sensitivity to a salty one. This change in the threshold for a salty stimulus is in the opposite direction to that for a sweet test stimulus (the adapting stimulus). Unfortunately, the evidence to date does not permit any generalizations about such cross-effects in adaptation.

Closely related are the phenomena of *successive contrast* in taste. An orange tastes sour after sugar, but sweet after lemon. Successive contrast is an adaptation phenomenon measured well above the absolute threshold.

Qualities of Taste. In contrast to odors, taste qualities are rather readily described in terms of four primaries: sweet, salty, sour, and bitter. The reader may object that the multiplicity of tastes does not augur well for such a reduction. But much of the richness of tastes comes not from gustation but from olfaction. If the nostrils are stopped up, or when the sense of smell is dulled by a cold, our world of tastes shrinks appreciably. Under such conditions, for example, we experience difficulty in distinguishing among the various kinds of meat, and beverages lose much of their zest and flavor. In addition to smell, stimulation of the various cutaneous receptors located in the mouth also contribute to our appreciation of food. For these reasons it is necessary to control or take account of nongustatory stimulation in searching for the basic taste qualities.

Although the mode of action of the stimuli is largely unknown, we do know, certainly better than in the case of smell, what types of stimuli give rise to these four primary qualities. The quality of sour is aroused by substances which contain acid ions. Thus, acetic acid makes vinegar sour, and citric acid is responsible for the sourness of lemons. The greater the concentration of acid ions, the more intense the sour taste tends to be.

The adequate stimulus for the pure quality of salt is sodium chloride (table salt). Other salts, such as the chlorides of ammonium, potassium, and calcium also taste salty, but there is an admixture of other taste qualities. The nitrates and sulfates also are salty, and it is probably true that the salty taste is aroused by free anions.

The specific chemistry of sweet and bitter stimuli is less well known. A large number of organic compounds, many of which are not ionized in solution, elicit sweet tastes. Many carbohydrates, especially sugar, taste sweet. Other compounds, such as saccharine, also do.

The bitter as well as the sweet taste is aroused by a variety of organic compounds. The alkaloids are especially bitter. Alkaloids are not the only stimuli for bitter, as is shown by the bitterness of magnesium sulfate (Epsom salts).

These four primary qualities are found in all kinds of blends and

fusions in gustatory experience. As with smell, the fusions are more or less easily analyzed into component qualities, depending upon the latter's degree of similarity. Further parallels to smell are compensations and contrast among taste qualities.

Physiological Basis of Taste Qualities. For each of the sense modalities, investigators have long sought to discover specialized receptors and their associated nerves responsible for the primary qualities of experience. The search has been moderately successful in most of the sensory systems. As far as gustation is concerned, the evidence is increasingly clear for the existence of separate receptors and nerve fibers for the taste qualities.

An invaluable technique in the investigation of this general problem is the isolation of single nerve fibers, and the recording of the impulses that pass along them. To isolate a single nerve fiber from a strand of nerve is a delicate operation, indeed, and the impulses must be highly amplified before they may be recorded. These techniques, rather recently perfected, have contributed greatly to our understanding of the physiology of the gustatory qualities.

In one outstanding investigation,[1] electrodes were placed on single fibers of the gustatory nerve of a cat. The animal's tongue was then stimulated by the four stimuli that correspond to the four primary qualities in man: sugar, quinine, salt, and acid. The nerve impulses recorded from single fibers conclusively demonstrated specialization among the receptors. Saline stimuli were found to activate only certain of the nerve fibers. Stimulation with quinine elicited impulses in other fibers. Acid also stimulated its own fibers, but, in addition, this stimulus activated the other two types as well. Hardly any impulses were detected when sugar was applied to the tongue. Although these relations are not as simple as we might hope, they establish, nevertheless, the fact that separate systems of fibers are present. It is the pattern of activity in these separate systems that provides the basis for differential response from one type of stimulus to another.

The following experiment[2] may be cited to illustrate the ways in which the physiological basis of quality may be inferred from psy-

[1] Pfaffman, C., Gustatory afferent impulses, *J. Cell & Comp. Physiol.*, 1941, *17*:243-258.

[2] Allen, F., and Weinberg, M., The gustatory sensory reflex, *Quart. J. Exper. Physiol.*, 1925, *15*:385-420.

chophysical data. The experiment utilized the fact that a rapid succession of discrete stimuli may be perceived as continuous. Below a certain frequency the stimuli are perceived as discrete. Above this frequency the stimuli are fused. The frequency at which fusion first occurs is called the *critical fusion frequency* (c.f.f.). To achieve

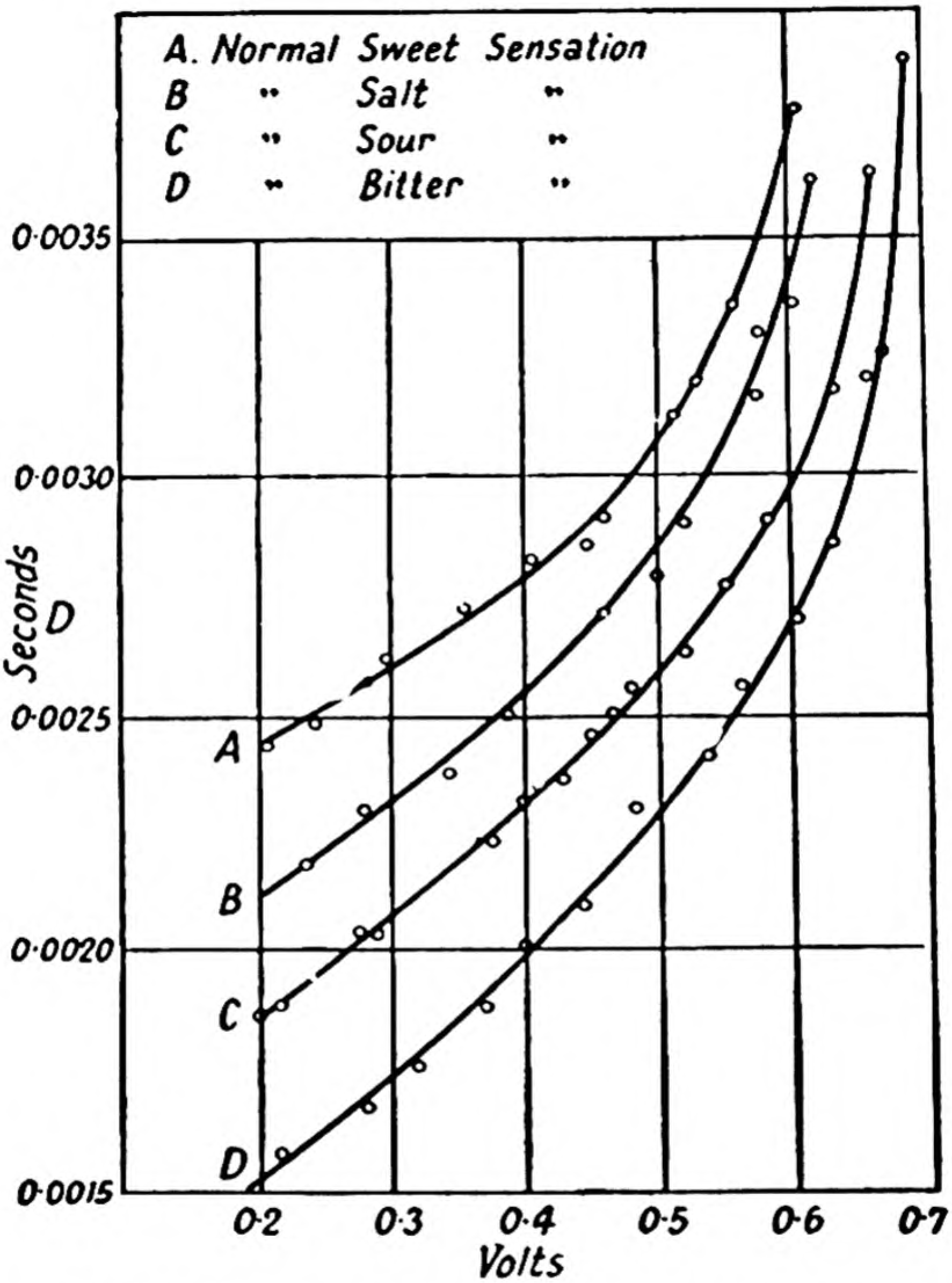


FIG. 26. Curves of Critical Fusion Frequency for the Four Basic Taste Qualities. (From R. W. Moncrieff, *The chemical senses,* 1946, p. 117, by permission of Leonard Hill, Ltd. After F. Allen and M. Weinberg, The gustatory sensory reflex, *Quart. J. Exper. Physiol.,* 1925, 15:390, by permission of the journal.)

rapid enough rates for fusion in the gustatory system, electrical stimulation had to be employed. This procedure was carried out at a number of different voltage levels. When c.f.f. was plotted against voltage level, it appeared not only that the fusion frequency decreased with voltage level, but also that the points clearly fell along four separate functions. The problem of identification of these functions remained. Since electrical stimulation gives rise to a sour taste, the function with the largest number of points was believed to represent the fusion frequencies for the sour receptors. The investigators then applied a drug to the tongue which is known to eliminate the taste for sweet and weaken the taste for bitter. When the fusion frequencies were again determined under this condition of partial taste anesthesia, only three functions were obtained, one of which showed only few points. The curve for sweet must have been absent, and the short curve represented bitter. In this manner all four curves were identified. Fig. 26 shows the four curves which are interpreted as presumptive evidence for four distinct kinds of sensory systems (note that the reciprocal of c.f.f. is plotted rather than c.f.f.).

We have reported this experiment in some detail because we feel that it is an excellent example of what can be achieved by analytic sensory research. In essence, these investigators combined several different techniques and types of data in arriving at a general, theoretical conclusion.

The Common Chemical Sense

We have already seen how taste and smell coöperate in the perception of food. There is, in addition, a third sensory system whose function is often closely associated with taste and smell. When we drink carbonated water, we experience a prickly feeling in the mouth. Onion and pepper have a sharp or irritating quality. These experiences are due in large measure to the *common chemical sense.*

Over the surface of the mouth and throat, on the tongue and in the nasal cavity, and in other parts of the body covered with a mucous membrane, there is a large number of free nerve endings which are sensitive to chemical irritants. It is this network of free nerve endings responsive to chemical irritants which makes up the common chemical sense.

The common chemical sense responds to a variety of stimuli.

Among the more usual ones are the spices, ammonia, chlorine, and other halogens. The severe discomfort that may attend exposure to certain war gases is due to the action of this sense.

We should examine some of the criteria which establish the common chemical sense as distinct from the olfactory and gustatory senses. There are, first of all, striking anatomical differences among the receptors of these three senses. Second, the distribution of these different types of cells over the body is quite different. It is characteristic of the receptors of the common chemical sense that they occur with considerable frequency in many different locations of the body. As mentioned previously, the receptors for smell and taste are each grouped in a rather restricted region of the body. Functional differences bear out the anatomical distinctions. The thresholds of the chemical sense tend to be much higher than taste thresholds, which, in turn, are higher than those for smell. Ethyl alcohol, for example, is an effective stimulus for all three senses. Compare the absolute thresholds:

0.44 percent per unit weight of air for smell.
14.00 percent per unit weight of water for taste.
25.00 percent per unit weight of water for the common chemical sense.

The respective qualities for this stimulus in each sense are: ethereal, sweet, and stinging. Finally, one or another of these three senses may be rendered nonfunctional, for example, by disease or anesthesia, without affecting the two remaining senses.

Experiment V

Olfactory and Gustatory Sensitivity

General Purpose. To demonstrate by a series of observations some of the functional properties of smell and taste.

Demonstration 1

Purpose. In this first demonstration we will observe the coöperation of smell and taste in the perception of flavors.

Materials. Pieces of potato, onion, carrot, cheese, and other food items are used. A blindfold serves to exclude visual cues. Dental cotton is required to stop up the nose in one part of the experiment.

Procedure. The subject is blindfolded, and his nose is carefully

stopped up with dental cotton. Thus, both visual and olfactory cues are excluded. A given substance is placed on the subject's tongue, and he is asked to report on its taste and, if possible, to identify it. Following this first report, the subject chews the food and again gives a report. The cotton is, then, removed, and the subject gives a fresh report on the taste and identity of the food. This procedure may be repeated for the various food substances.

Demonstration 2

Purpose. To show adaptation to an olfactory stimulus even though it is not perceived because it is being masked by another stimulus.

Materials. Peruvian balsam and eau de Cologne are the stimuli. Two small receptacles are used to present the stimuli.

Procedure. The subject is blindfolded and first asked to sniff Peruvian balsam in order to acquaint him with its odor. He then sniffs for about 2 minutes a mixture of Peruvian balsam and eau de Cologne. This mixture has been prepared so that the eau de Cologne masks the balsam. After the 2 minutes are up, the subject immediately sniffs the balsam again. What has happened to the intensity of the odor of Peruvian balsam?

Demonstration 3

Purpose. To show how the ability to detect the components in a mixture may depend upon the relative degree of adaptation to the separate components.

Materials. This time eau de Cologne and camphor are the stimuli. A blindfold is required.

Procedure. The subject tries to identify the components of a mixture of eau de Cologne and camphor. This mixture has been prepared so that the eau de Cologne nearly or completely masks the camphor. After sniffing the mixture, the subject is presented with just eau de Cologne, and he adapts to this stimulus for about 2 minutes. He again sniffs the original mixture of eau de Cologne and camphor and attempts to identify the components. Is he more successful this time?

General References

Boring, E. G., Langfeld, S. H., and Weld, H. P., *Psychology: a factual textbook*, New York: John Wiley and Sons, 1935, Chap. 6.

Boring, E. G., Langfeld, S. H., and Weld, H. P., *Foundations of psychology*, New York: John Wiley and Sons, 1948, Chap. 15.

Crozier, W. J., Chemoreception. In C. Murchison (ed.), *A handbook of general experimental psychology,* Worcester: Clark University Press, 1934.

Moncrieff, R. W., *The chemical senses,* New York: John Wiley and Sons, 1946.

Parker, G. H., *Smell, taste, and allied senses in vertebrates,* Philadelphia: J. B. Lippincott Co., 1922.

Woodworth, R. S., *Experimental psychology,* New York: Henry Holt and Co., 1938, Chap. 20.

Smell

Dimmick, F. L., The investigation of the olfactory qualities, *Psychol. Rev.,* 1927, *34*:321-335.

Elsberg, C. A., The sense of smell: VIII. Olfactory fatigue, *Bull. Neurol. Inst. N. Y.,* 1935, *4*:479-495.

Elsberg, C. A., Olfactory tests. In O. Glasser (ed.), *Medical physics,* Chicago: Year Book Publications, 1944.

Jerome, E. A., Olfactory thresholds measured in terms of stimulus pressure and volume, *Arch. Psychol.,* 1942, No. 274.

Zwaardemaker, H., *L'odorat,* Paris: Doin, 1925.

Taste

Allen, F., and Weinberg, M., The gustatory sensory reflex, *Quart. J. Exper. Physiol.,* 1925, *15*:385-420.

Holway, A. H., and Hurvich, L., Differential gustatory sensitivity to salt, *Amer. J. Psychol.,* 1937, *49*:37-48.

Pfaffman, C., Gustatory afferent impulses, *J. Cell. Comp. & Physiol.,* 1941, *17*:243-258.

Richter, C. P., Total self-regulatory functions in animals and human beings, *Harvey Lect.,* 1943, *38*:63-103.

Richter, C. P., and MacLean, A., Salt taste thresholds of humans, *Amer. J. Physiol.,* 1939, *126*:1-6.

Common Chemical Sense

Crozier, W. J., Regarding the existence of the "common chemical sense" in vertebrates, *J. Comp. Neurol.,* 1916, *26*:1-8.

»»» 6 «««

VISION

MAN'S acquaintance with the world in which he lives depends to a considerable extent on his sense of sight. Orientation in space, discrimination and manipulation of colors, forms, and objects—all these and many other activities are critically dependent on the proper functioning of the visual system. Responses to visual stimuli are closely integrated and coördinated with other response mechanisms such as the auditory, static, and motor systems of the organism. In this chapter we shall examine the basic visual functions, the properties of visual stimuli and their measurement, the characteristics of visual experience and discrimination. A thorough analysis of these basic functions is indispensable for an understanding of man's rich and varied visual experience and of the role it plays in behavior.

The Visual Stimulus

Radiant Energy. Visual experience depends upon the stimulation of the sensitive receptors of the eye by radiant energy. Although the physicist's conception of the nature of radiant energy has undergone many changes in recent years, we may still think for our purposes of radiant energy as wave motion with two basic dimensions: wave length and intensity.

Wave Length. The wave lengths of electromagnetic motion extend over an enormous range. Some of them are exceedingly short and to speak conveniently about them, a special unit of measurement—the millimicron ($m\mu$)—is employed. A micron is one-millionth of a meter, and a millimicron is a thousandth of a micron. Wave lengths of radiant energy as short as 1/1000 of a millimicron and as long as 400 billion millimicrons are known to exist. Along this continuum of waves, *visible* radiation, or *light*, forms an extremely nar-

row band. This visible spectrum extends from about 400 to about 760 millimicrons.[1]

Intensity. The intensity of an electromagnetic wave can, of course, be measured in strictly physical terms. The intensity of light was, however, originally measured in terms of its effect on the human eye and only later did physicists specify the intensity of light in strictly physical terms, such as ergs or watts. As a result, two systems of measuring the intensity of light have come into practice, and these must be carefully distinguished: the physical or radiometric and the psychophysical or photometric units. Physical measurement is made by light-sensitive instruments, independently of the response of the eye to light, and results in such units as ergs or watts. Photometric measures, on the other hand, are based on the judgments of human observers under standardized conditions. The photometric measures are known as units of *brightness.*

The photometric units of brightness are defined in reference to the following idealized standard conditions. A source of light, known as the standard candle, is placed at the center of a sphere, the radius of which is 1 meter. It is assumed that this sphere diffuses *all* of the light which it receives from the candle at the center. The sphere is illuminated from the inside and the *meter-candle* is the unit of illumination under these conditions. The total amount of light falling on 1 square meter of the total surface of the sphere is called a *lumen.* Now let us consider the light diffused into the space beyond the sphere by 1 square meter. Since the sphere transmits all the light it receives, the brightness of the sphere as viewed from the outside is 1 *lumen per square meter.* It would have been possible to use lumen per square meter as a unit of brightness. However, it would be too small for convenience, and a unit ten times as large is typically used. This more convenient unit of brightness is called the *millilambert* A millilambert, then, is the brightness of a perfectly diffusing sphere, with 1 meter radius and 10 standard candles at its center Other sources of light of unknown intensity can be calibrated by matching them in brightness to the standard sphere The device usually employed for this purpose is called the *photometer.*

[1] As ordinarily viewed, the spectrum extends to about 760 mμ. Recent determinations have shown that wave lengths as long as about 1000 mμ can be seen if the energy is made great enough.

A convenient instrument for making a photometric match is the *Macbeth illuminometer* In this illuminometer, the surface whose brightness is to be measured appears as a circle surrounded by an illuminated surface which forms a ring around this circle. The brightness of this annular ring can be adjusted until the boundary between the circle and the ring can no longer be discriminated. Since the brightness of the annular ring is known, the value of the test surface can then be specified. The annular ring used for the brightness matches is white. Sometimes it is necessary to determine the brightness of a colored surface. It is not then possible to make the boundary between circle and ring disappear completely. For this kind of match, called *heterochromatic* photometry, it is necessary that the observer make his judgment disregarding the difference in hue as well as he can.

In the study of visual processes, we would ideally want to measure the intensity of the light as it strikes the retina. If the pupil of the eye never varied in size, the photometric brightness of a given surface would be directly related to the intensity of the light impinging on the receptor cells of the retina. The size of the pupil, however, varies with the intensity of the light. In order to take the size of the pupil into account, a unit called the *photon* is frequently used. When the brightness of a surface is one-tenth of a millilambert and the effective pupillary area is one square millimeter, the value of the stimulus is one photon.

Homogeneous and Heterogeneous Light. Most visual stimuli consist of many different wave lengths of light. They are *heterogeneous* with respect to wave length. By passing a beam of heterogeneous light through a prism, it is possible to break it up into its components, i.e., into *homogeneous* wave lengths. These homogeneous wave lengths are used to investigate the laws of stimulus mixture.

The Dimensions of Color

The Attributes of Color. The psychophysics of color is concerned with the establishment of relations between the physical characteristics of the stimulus and the attributes of visual experience. One of the most fundamental distinctions which we can make in studying the properties of color is that between *chromatic* and *achromatic* colors. The chromatic colors are the rainbow colors: the

reds, the yellows, the greens, and blues, and the whole manifold of colors which shade into each other, from the reddest red to the bluest blue. This mode of variation of colors defines the dimension of *hue.*

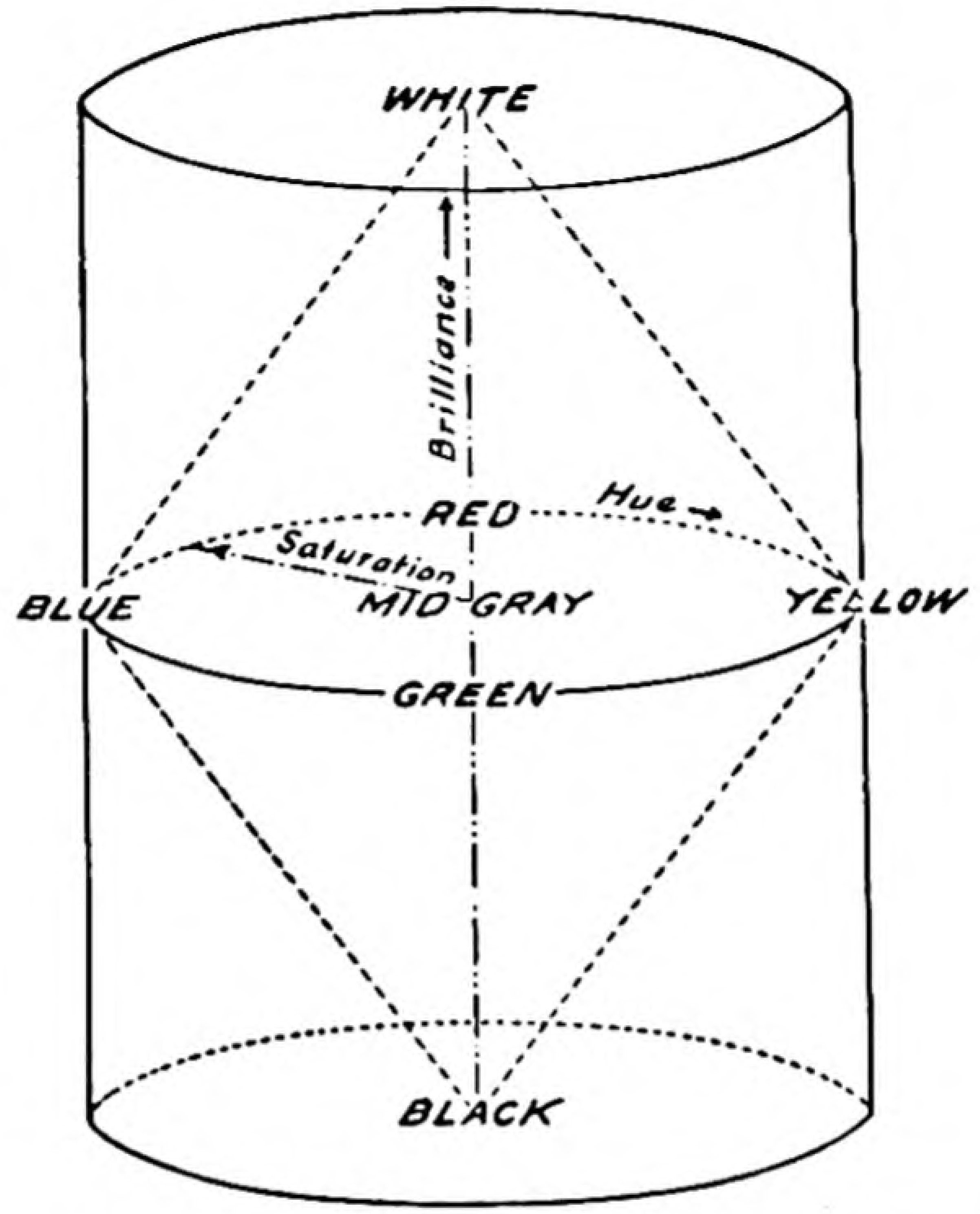


FIG. 27. The Psychological Color Solid. The vertical axis represents the dimension of brilliance, the circumference represents hue, and the lateral distance from the vertical axis, saturation. (From L. T. Troland, *Principles of psychophysiology, Vol. I.*, 1930, p. 251, by permission of D. Van Nostrand Company, Inc.)

The achromatic colors comprise the series from the blackest black through the grays to the whitest white. This series from black to white parallels the dimension of visual experience called *brilliance.* These two dimensions, hue and brilliance, are not sufficient for a complete description of the properties of color. Let us look at a red surface. It may be full and rich in redness or it may be only slightly

tinted with red, yet in both cases we describe it as *red.* Any hue can thus vary in *saturation.* Saturation of a chromatic color refers to the difference between it and an achromatic color of the same brilliance.

It is necessary, then, to specify three dimensions of variation—hue, brilliance, and saturation—for the full description of the properties of a color. It is helpful in the description of color to think of these three dimensions as arranged in the form of a solid figure. Such a color solid is shown in Fig. 27. The vertical axis represents the dimension of brilliance. The circumference of the solid represents hue, and the lateral distance from the vertical axis indicates degree of saturation. As the dotted lines in Fig. 27 indicate, dark and brilliant colors have a smaller range of saturation than those intermediate in brilliance. It is possible to ascribe a location in this solid to any given color.

Having defined the physical characteristics of the visual stimulus and the attributes of colors, our next task is to examine the functional relationships between them. Variations in any one of the three attributes of color are a *joint function* of both the wave length and the intensity of the stimulus.

Hue and Wave Length

We remember that the visible spectrum extends from about 400 to about 760 millimicrons. With intensity held constant at a medium level, variations in wave length over the visible range lead to changes in hue, covering the whole gamut of colors from violet to red. Continuous variation in wave lengths generates the rich manifold of spectral colors. This spectrum is produced when a prism breaks down heterogeneous light into its component homogeneous wave lengths. The long wave lengths correspond to reddish hues. As the wave lengths become shorter, the colors pass through oranges to yellow (575 mμ). As the wave lengths become even shorter, the yellows become more greenish and, finally, a clear green appears at about 505 mμ. As we approach yellow from the long wave lengths, the reddish aspect of the hues becomes less prominent and as we pass yellow, the greenish aspect becomes more noticeable. Since yellow sharply divides the reds from the greens, it is considered a *primary color.* In a similar way, as we further shorten the wave length beyond a clear green, the yellow tint disappears and the

greens become more and more bluish. Clear blue is, then, seen at about 480 mμ. Like yellow, green and blue are primary colors. A clear red is also a primary color but it does not appear as such in the spectrum. To obtain such a primary red, we must mix long and short wave lengths. Primary red is thus produced only by heterogeneous light.

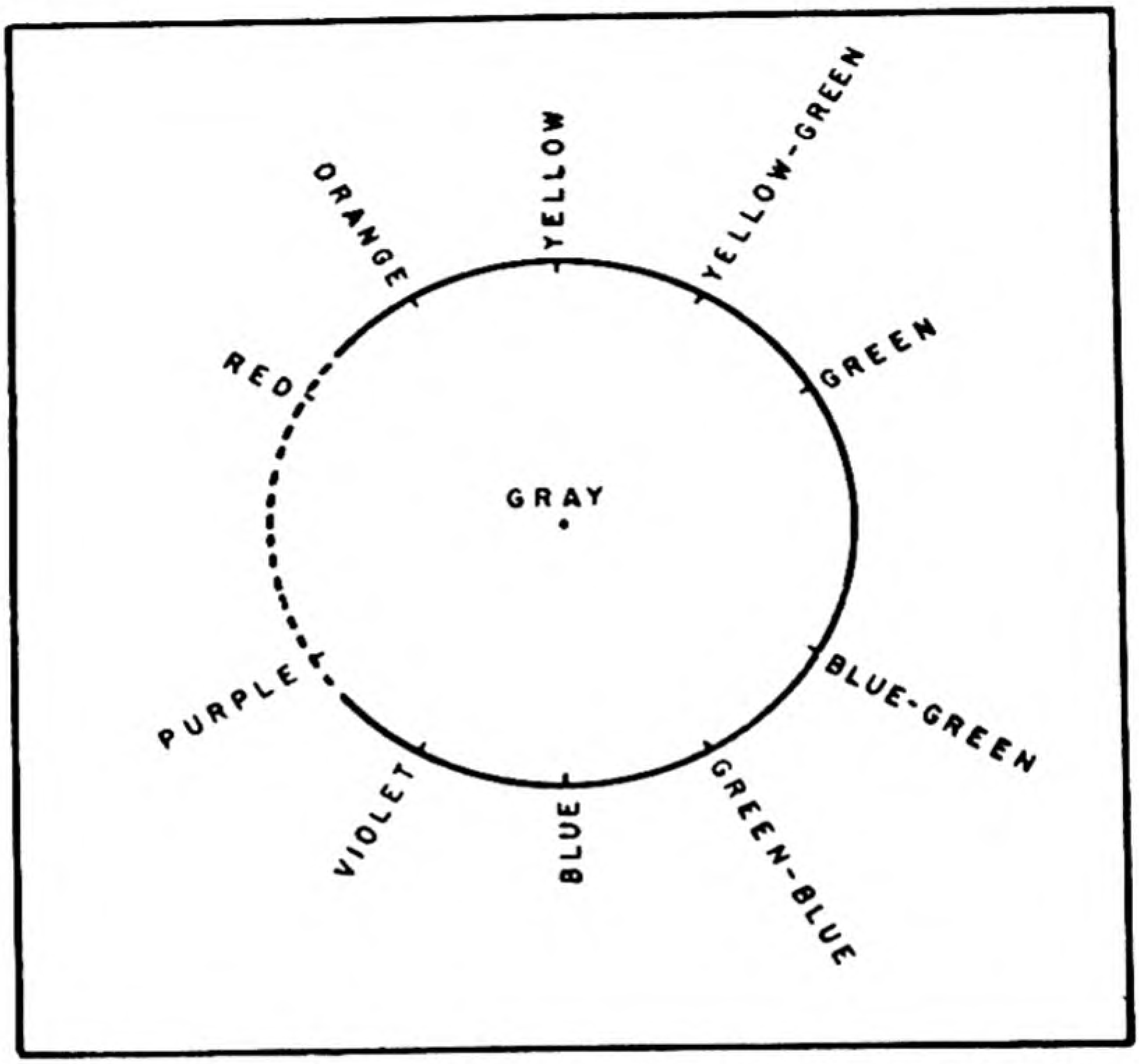


Fig. 28. The Color Circle. The spectral colors are represented by the solid part of the circumference, the dotted part stands for colors which can be obtained by mixing spectral lights.

The very short wave lengths (beyond 480 mμ) are seen as violet. The violets take on a slight reddish aspect. Thus, the spectrum nearly completes a circle from an orange-like red around to a slightly reddish blue. The circle, however, is not complete. It is represented by the solid part of the circumference in Fig. 28. The circle may be completed by proper mixtures of long and short wave lengths yielding the purples and a primary red (dotted part of Fig. 28). It is this circle which also appears as the circumference of the color solid in Fig. 27.

Hue and Intensity

Hue is also a function of the intensity of the stimulus. As intensity is decreased, the spectrum becomes less and less saturated, i.e., the hues become less and less pronounced. At low intensities the spectrum becomes achromatic, and there is no differentiation of wave lengths with respect to hue. Such weak lights are visible even though hue is absent. For a given homogeneous wave length, this interval between the absolute threshold of visibility and the first appearance of hue is the *photochromatic interval.* The photochromatic interval is very small or zero for the very long wave lengths (above about 665 mμ). The photochromatic interval can best be demonstrated by stimulating the peripheral part of the retina. The rods, which predominate over the cones in the periphery, have a much lower threshold than the cones and do not function in chromatic vision.

Even when the intensity is high enough to produce a spectrum of hues, hue still varies with intensity. This change may be investigated by requiring a subject to match, with respect to hue, two lights which differ in intensity. The wave length of one of the lights is kept constant, and the subject varies the wave length of the other. The results of such an experiment appear in Fig. 29. For any given match, the standard stimulus had a fixed wave length and a fixed intensity. The comparison or variable stimulus had another intensity, and the subject adjusted its wave length until its hue matched that of the standard. The differences in wave length between standard and variable required for a hue match are plotted in Fig. 29. As the figure shows, no adjustment of wave length was required for the three primaries, blue, green and yellow. Hue is invariant with intensity at these points. For other wave lengths, however, the hue shifts with an increase in intensity. As the arrows indicate, these shifts are in the direction of blue and yellow as the intensity increases. A violet at, say, 450 mμ, appears more blue as the intensity is increased. Similarly, a greenish blue takes on more of a blue tint with increased intensity. At the other end of the spectrum, an orange becomes more yellow as does a greenish yellow. This change in hue as a function of intensity is known as the *Bezold-Brücke phenomenon.* Although the phenomenon is experimentally established, the

changes are not very great and often almost within the range of the differential threshold for hue.

Brilliance and Intensity

At a given wave length. a minimum intensity or energy is required for visual response. This minimum intensity is the absolute threshold at that wave length. Increases in intensity beyond the absolute

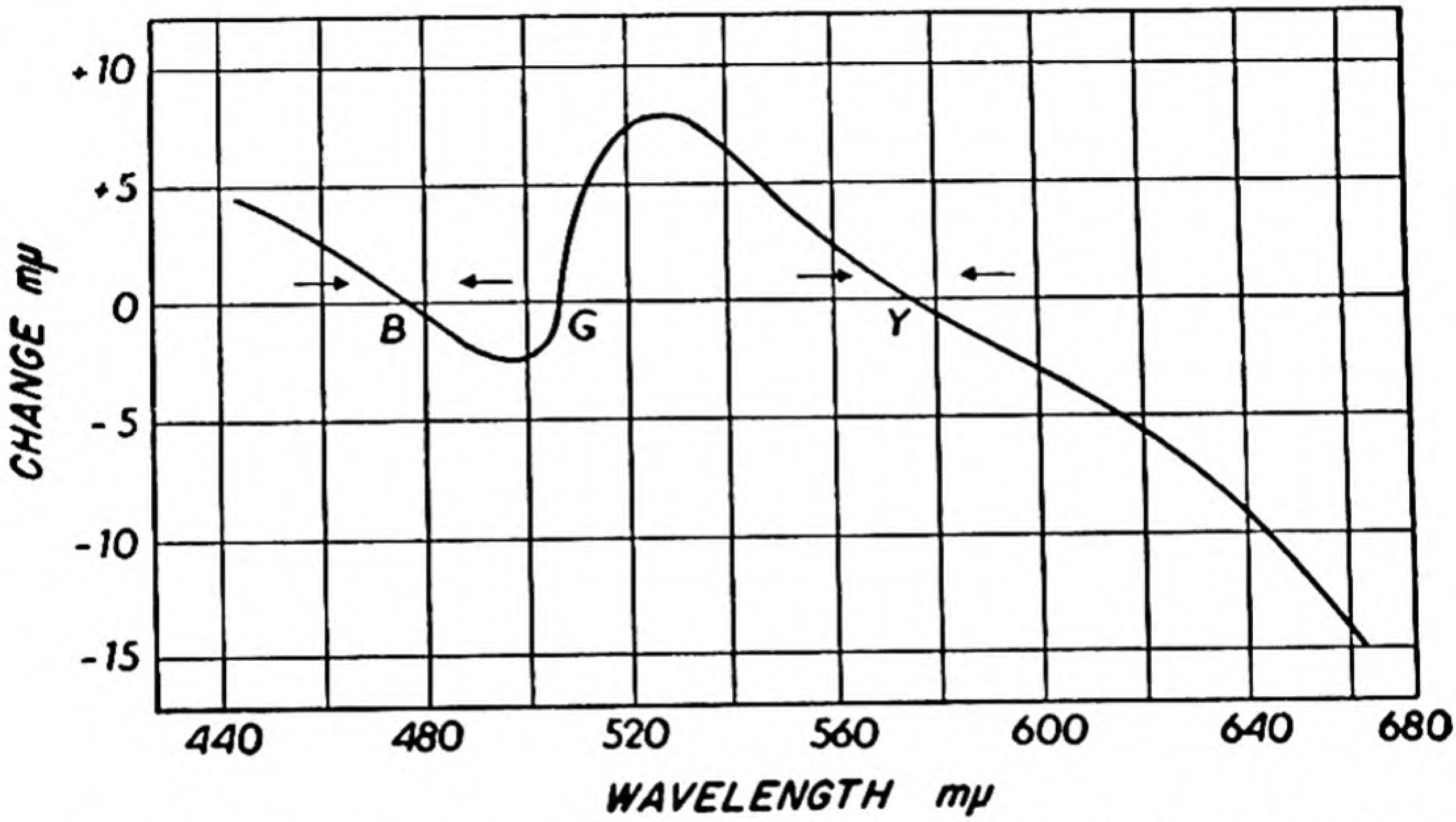


FIG. 29. The Bezold-Brücke Phenomenon: Changes in Hue as a Function of Intensity. The graph shows by how many millimicrons a stimulus of given wave length had to be altered in order to match a standard of the same wave length but of smaller intensity. Note the three invariant points—*B*, *G*, and *Y*. (From L. T. Troland, *Principles of psychophysiology, Vol. II*, 1930, p. 170, by permission of D. Van Nostrand Company, Inc.)

threshold result in increases in brilliance. The relationship between intensity and brilliance is not, however, a simple one. The relationship obtained depends on the state of adaptation of the retina, on the area of the retina stimulated, the wave length of light, etc. The general nature of the relationship for a controlled set of conditions appears in Fig. 30, where brilliance (number of just noticeable differences above threshold) is plotted against the logarithm of the intensity.[2] At the low end of the scale a given change in the loga-

[2] The logarithm of intensity is used principally to condense the extremely wide range of intensities used.

rithm of intensity results in a smaller number of just noticeable differences of brilliance than in the middle or upper part of the scale.

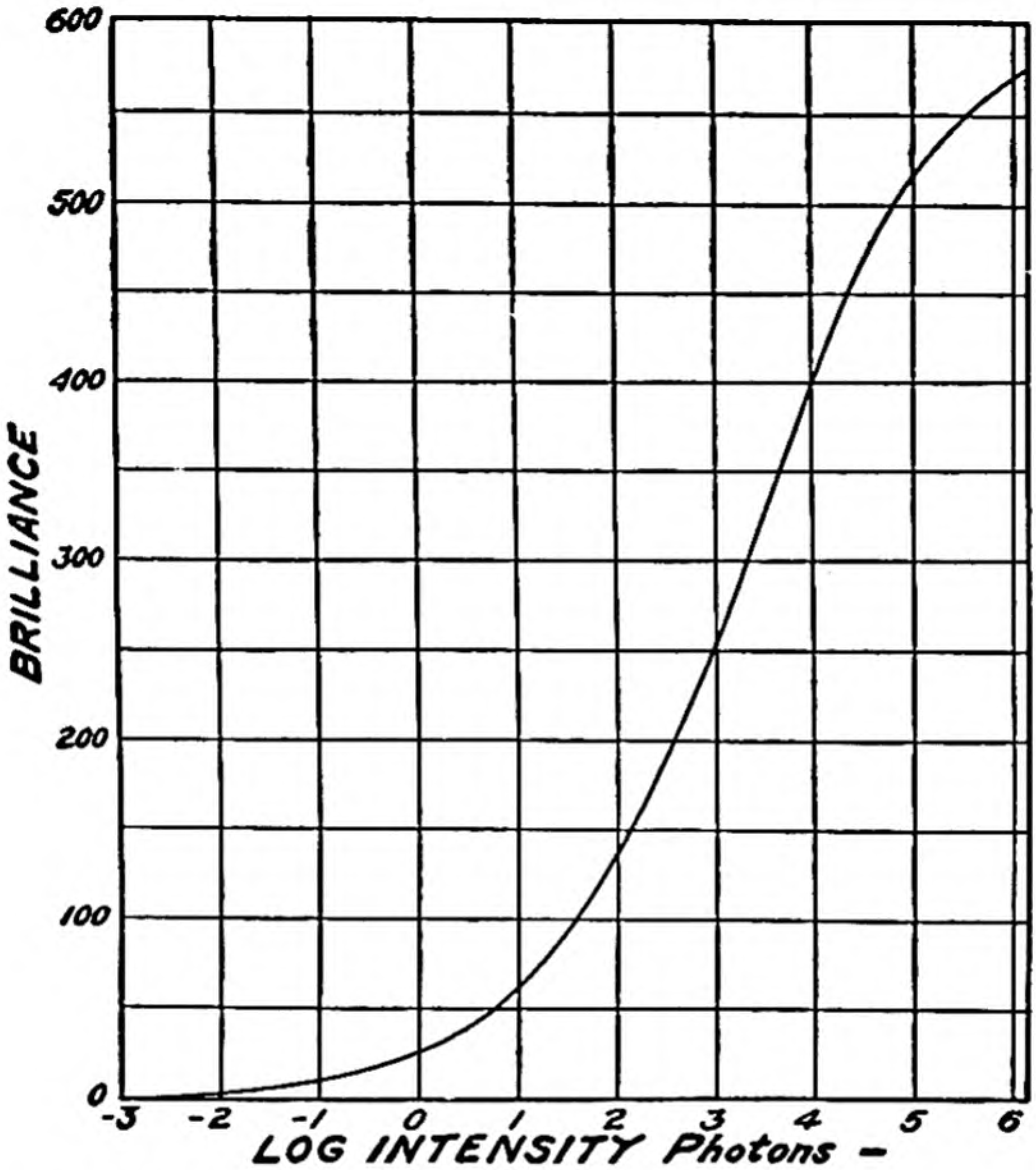


FIG. 30. Brilliance as a Function of the Intensity of the Stimulus. This curve shows how brilliance grows with the logarithm of stimulus intensity. The brilliance scale was obtained by summation of just noticeable differences in brilliance above the absolute threshold. (From L. T. Troland, *Principles of psychophysiology, Vol. II,* 1930, p. 77, by permission of D. Van Nostrand Company, Inc.)

The function shown in Fig. 31 was obtained by determining the differential threshold for intensity at successive levels of intensity. At any level of intensity the differential threshold may be expressed relative to the intensity at which the threshold is determined, i.e.,

as $\triangle I/I$. When $\triangle I/I$ is plotted as a function of the logarithm of intensity, the function in Fig. 31 is obtained. As was found for $\triangle I/I$ in the discrimination of auditory intensity, the curve drops rapidly and then levels off over a wide range of intensity. In the middle range of intensity $\triangle I/I$ drops to a minimum.

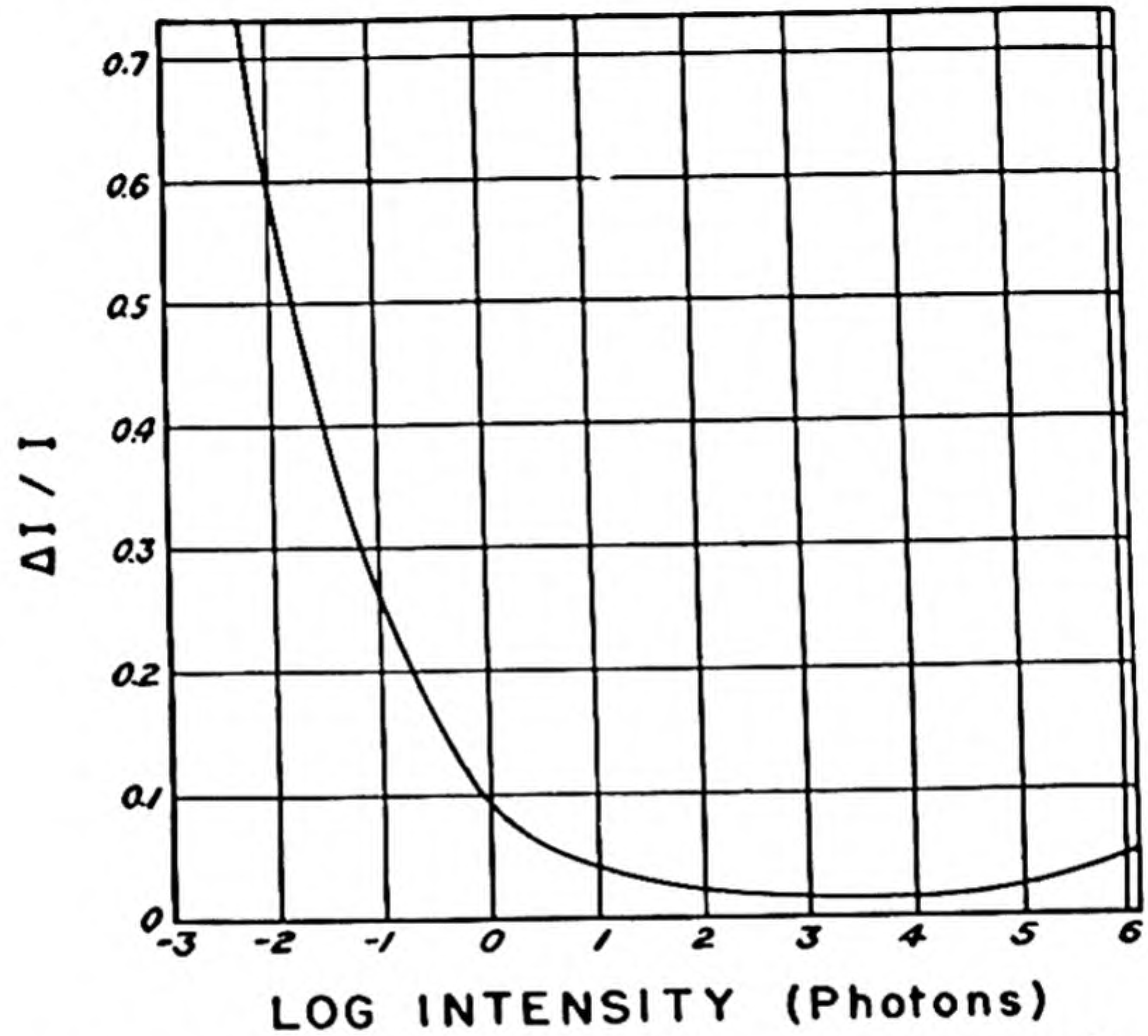


FIG. 31. $\triangle I/I$ as a Function of Intensity. This curve shows how relative differential sensitivity to the intensity of the visual stimulus varies with the logarithm of the intensity of the stimulus. At any given level, the differential threshold is expressed relative to the intensity at which the determination is made, i.e., in terms of $\triangle I/I$. (From L. T. Troland, *Principles of psychophysiology, Vol. II*, 1930, p. 78, by permission of D. Van Nostrand Company, Inc.)

Brilliance and Wave Length: Visibility Curves

Brilliance varies as a function of wave length. When the amount of radiant energy is held constant, different wave lengths do not produce the same amount of brilliance. The dependence of brilliance on wave length is usually shown by determining the relative amounts

of energy required for different wave lengths of light to evoke constant brilliance.

A white light of fixed intensity (and brilliance) is used as a standard. The subject adjusts the intensity of a number of stimuli of different wave lengths until each matches the standard in brilliance (heterochromatic photometry). The wave lengths of the homogeneous stimuli are selected so as to cover the range of the visible spectrum. Different energy values are required for the various wave

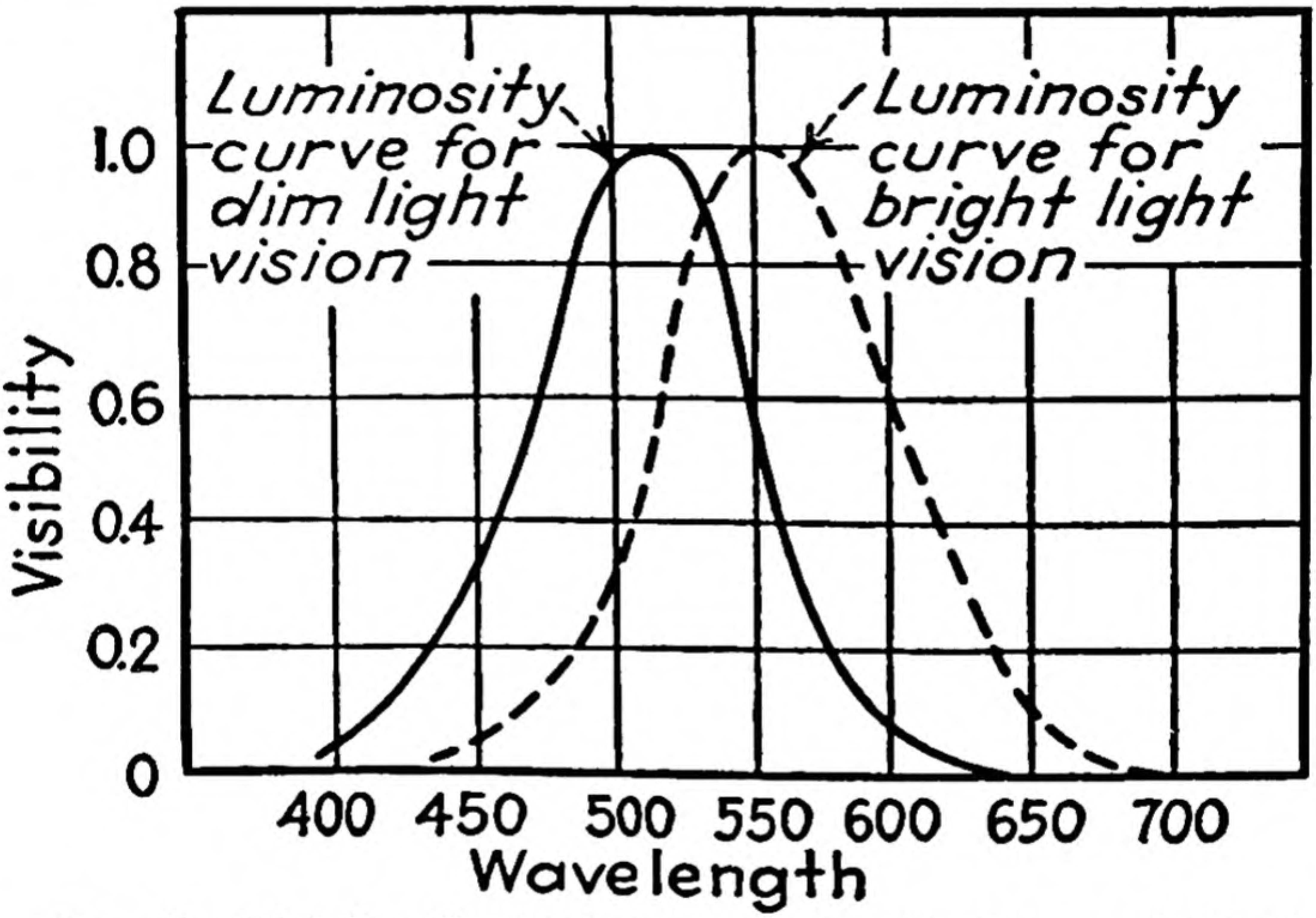


FIG. 32. Visibility (luminosity) Curves for Dim-Light and Bright-Light Vision. (By permission from *Recent experiments in psychology* by L. W. Crafts, *et al.*, p. 139, copyrighted 1938, McGraw-Hill Book Co., Inc.)

lengths to match the brilliance of the standard. If we compute the reciprocals of the energy values, we obtain a measure of the relative visibility of the different stimuli. Another procedure calls for the determination of the absolute thresholds (minimum energy required for detection) with stimuli of different wave lengths. Again, the reciprocal of the energy required for detection is a measure of rela-

tive visibility. To express the results of the two procedures along a common scale, values of visibility are expressed relative to the most visible stimulus. The visibility value of the most visible stimulus, then, is 1.0, and the other stimuli are designated by fractions less than 1.0.

The visibility values obtained by these procedures are shown in Fig. 32. Two distinct curves result: one of them obtained from matches with a bright light (bright-light vision), and the other from the relative values of absolute sensitivity (dim-light vision). The visibility curve for bright-light vision is called the *photopic* curve; that for dim-light vision is called the *scotopic* curve.

The relation of visibility to wave length is affected importantly by the level of illumination. In bright-light vision, maximum visibility occurs at about 555 mμ. At very weak levels of stimulation, the point of maximum sensitivity shifts toward the short end of the spectrum and occurs at about 511 mμ. Under dim-light conditions, the whole visibility curve is displaced toward the short end: relative sensitivity to short wave lengths increases while relative sensitivity to the long wave lengths decreases.[3]

The difference between the two curves is due to the fact that it is the cones which principally function in bright-light vision, and the rods in dim-light vision. These two types of receptors differ in their relative sensitivity to stimuli of different wave lengths. The relative sensitivity of the cones is responsible for the photopic curve; the relative sensitivity of the rods determines the scotopic curve. One bit of evidence for the latter statement is that with different intensities of light, the shift in sensitivity does not occur when the stimulus is restricted to the fovea (or rod-free area). Furthermore, in the case of certain animals, which possess predominantly rods and few cones, the shift does not occur.

Impressive support for this explanation of the differences between photopic and scotopic curves comes from the photochemistry of the retina. A chemical substance called *rhodopsin* can be extracted from the rods but not from the cones. The substance absorbs light and is thereby broken down. The rate of absorption depends on the wave length of the radiation. It has been found that the absorption curve

[3] This shift in relative sensitivity is often called the Purkinje effect, after the physiologist who first discovered it.

of rhodopsin is almost the same as the scotopic visibility curve. Fig. 33 demonstrates the close correspondence of the scotopic curve and the absorption curve of rhodopsin. The photochemistry of the action of the cones is still largely unknown.

Saturation and Wave Length

The different wave lengths of the spectrum are not equally sat-

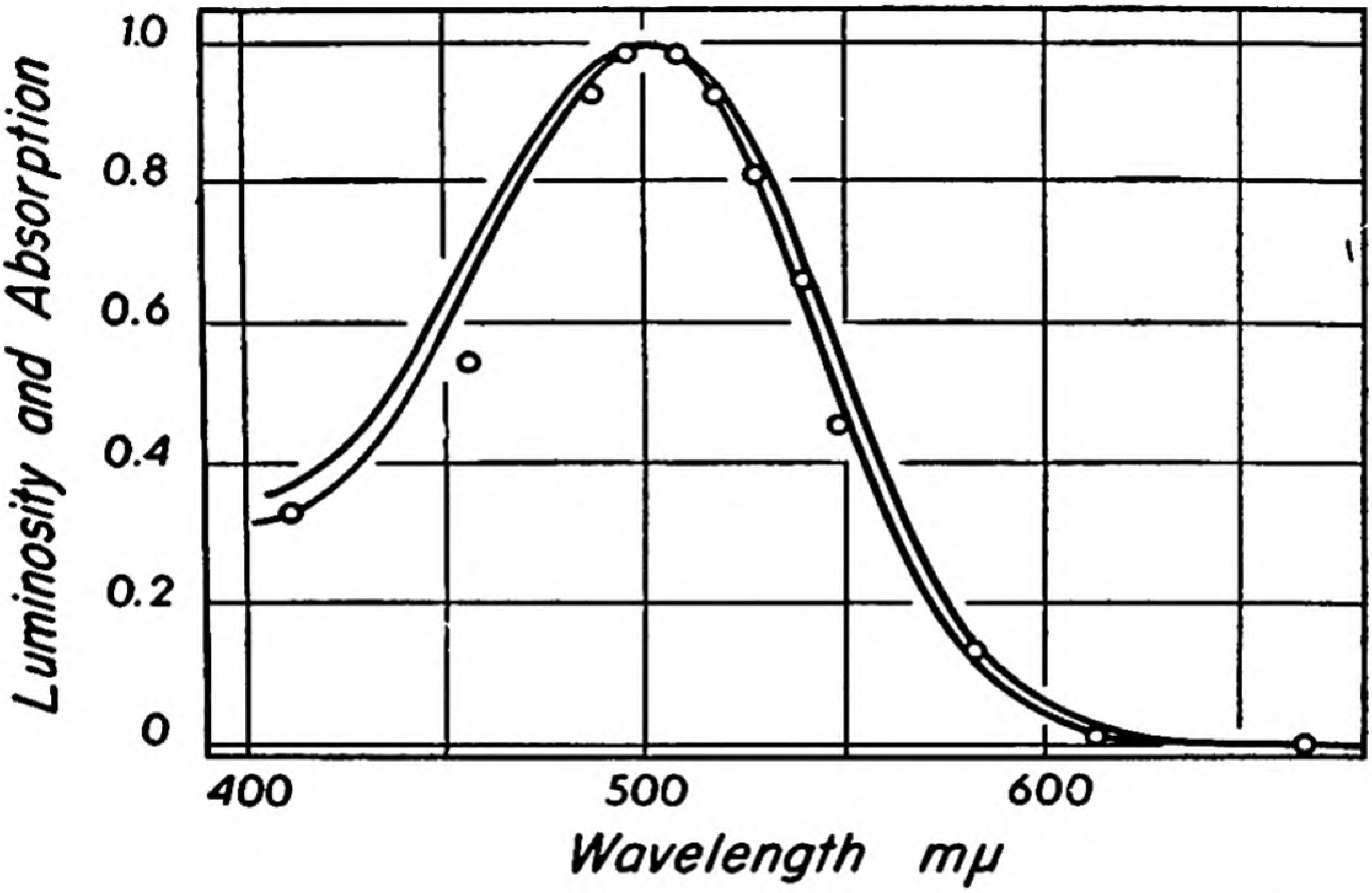


FIG. 33. Correspondence of the Scotopic Visibility Curve and the Absorption Curve of Rhodopsin (circles). (From S. Hecht, S. Shlaer and M. H. Pirenne, Energy, quanta, and vision, *J. Gen. Physiol.*, 1942, *25*:831, by permission of the journal and authors.)

urated. Turning back to the color solid in Fig. 27, we remember that saturation is degree of difference of a chromatic from an achromatic color. Spectral yellow is less different from white than is red or blue.

In one experiment this distance was measured in terms of the number of just noticeable differences in saturation between the spectral color and white. Changes in saturation were obtained by adding varying amounts of white light to the homogeneous spectral color. A scale of saturation was constructed by determining the

amounts of white light which had to be added to obtain successive just noticeable differences in saturation. This procedure was followed for each of several spectral colors. The number of steps required completely to desaturate each of these various spectral colors is shown in Fig. 34. There we can see that yellow (575 mμ) requires the smallest number of steps, whereas blue and red require

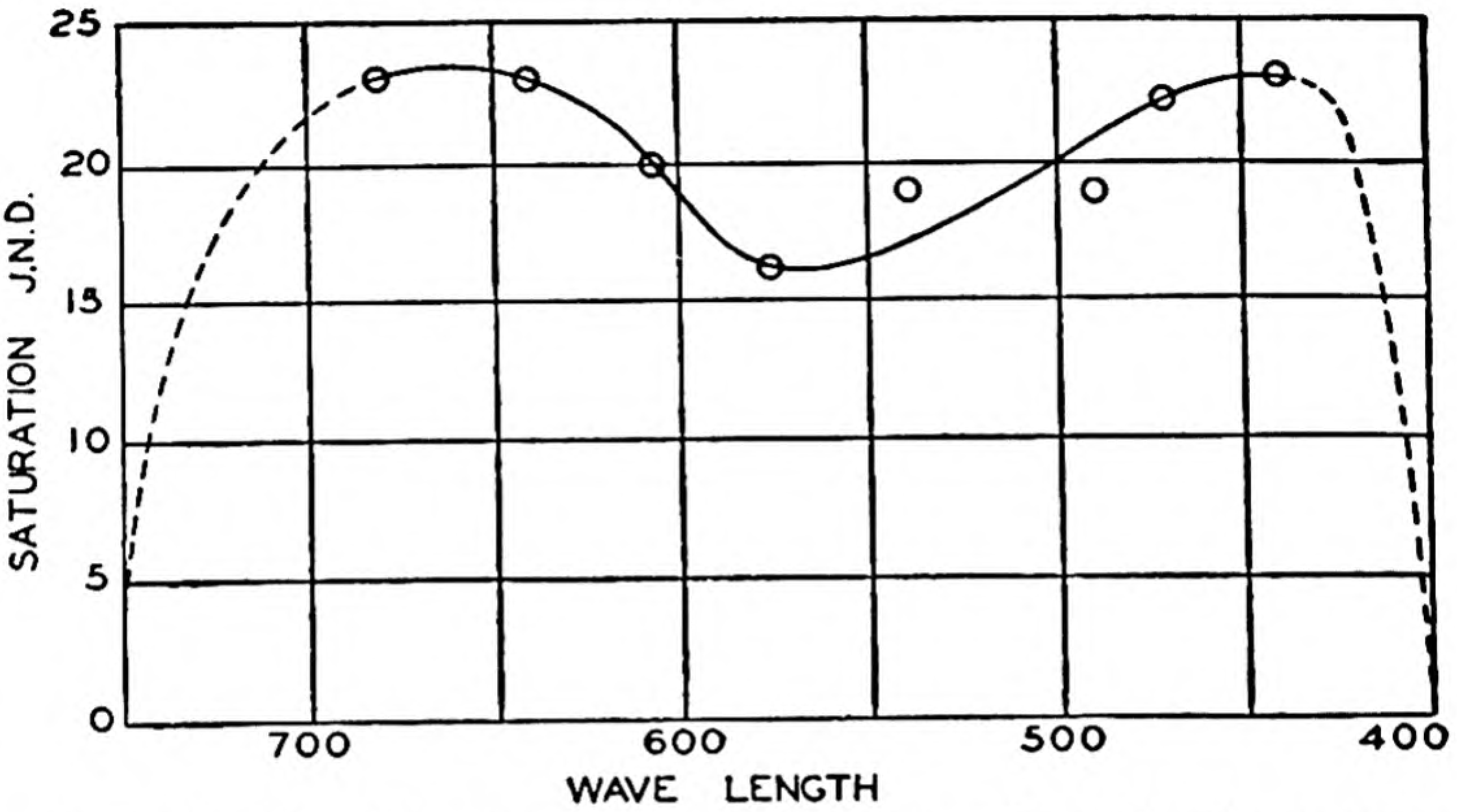


FIG. 34. The Saturation of Spectral Colors. The scale of saturation was constructed by determining the amounts of white light which had to be added to the spectral color in order to obtain successive just noticeable differences in saturation. The points plotted as a function of wave length represent the number of steps required for complete desaturation. (From E. G. Boring, The psychophysics of color tolerance, *Amer. J. Psychol.*, 1939, 52:393, by permission of the journal and author.)

the largest number. The spectrum, then, is maximally saturated at its extremes and minimally at yellow.

Saturation and Intensity

A color of very low intensity usually has no saturation at all. It will be remembered that all but the very long wave lengths will evoke a visual response before they appear chromatic. During this photochromatic interval, such wave lengths have zero saturation. As the threshold of hue is passed, saturation increases as a function

of intensity, reaches a maximum in the middle range of intensities, and then declines as the intensities become very high. For each wave length there is a different intensity yielding maximum saturation. A highly saturated color reaches maximum saturation at a lower intensity than a less saturated color.

Stimulus Mixture

Two-Component Mixture. The hues of the spectrum may be obtained by passing a beam of white light through a prism and breaking it up (by refraction) into its component wave lengths. This is not the only way, however, in which the hues of the spectrum may be obtained. They may also be produced by mixing homogeneous rays of light. For example, a homogeneous red light mixed with a homogeneous yellow light is seen as orange which matches a spectral orange. We may note from this example that the hue of the mixture lies on the color circle between those of the two homogeneous components. The mixture will, moreover, be less saturated than the more saturated of the components.

Let us keep the wave length of the red component fixed and vary that of the yellow component toward green. As we do so, the hue of the mixture moves from orange toward yellow and becomes less and less saturated. If we further change the wave length of our variable component a little toward blue, the mixture becomes achromatic (gray). The homogeneous red and homogeneous bluish green are said to be *complementaries*. The existence of complementaries could also have been demonstrated by choosing as our components yellow and blue, yellow-green and violet, or, indeed, any two components which lie opposite each other on the color circle. From a great deal of experimental work on stimulus mixture, the following laws have been formulated.

Law of Intermediates. The hue of the mixture of any two homogeneous components is intermediate on the color circle between the hues of those components. The hue of the mixture is closer to that component which has the greater intensity. The mixture is less saturated than one or both of the components.

Law of Complementaries. If two hues are diametrically opposite each other on the color circle, the mixture of their stimuli in the proper intensity ratio appears gray.

Many of the pairs of stimuli which are complementary are homogeneous lights, e.g., spectral yellow and spectral blue. Since the circle of homogeneous spectral colors is not closed, some homogeneous spectral lights do not have complementaries which are also homogeneous lights. For example, a green homogeneous light has no homogeneous complementary but must be mixed with a mixture of red and blue lights to produce a gray.

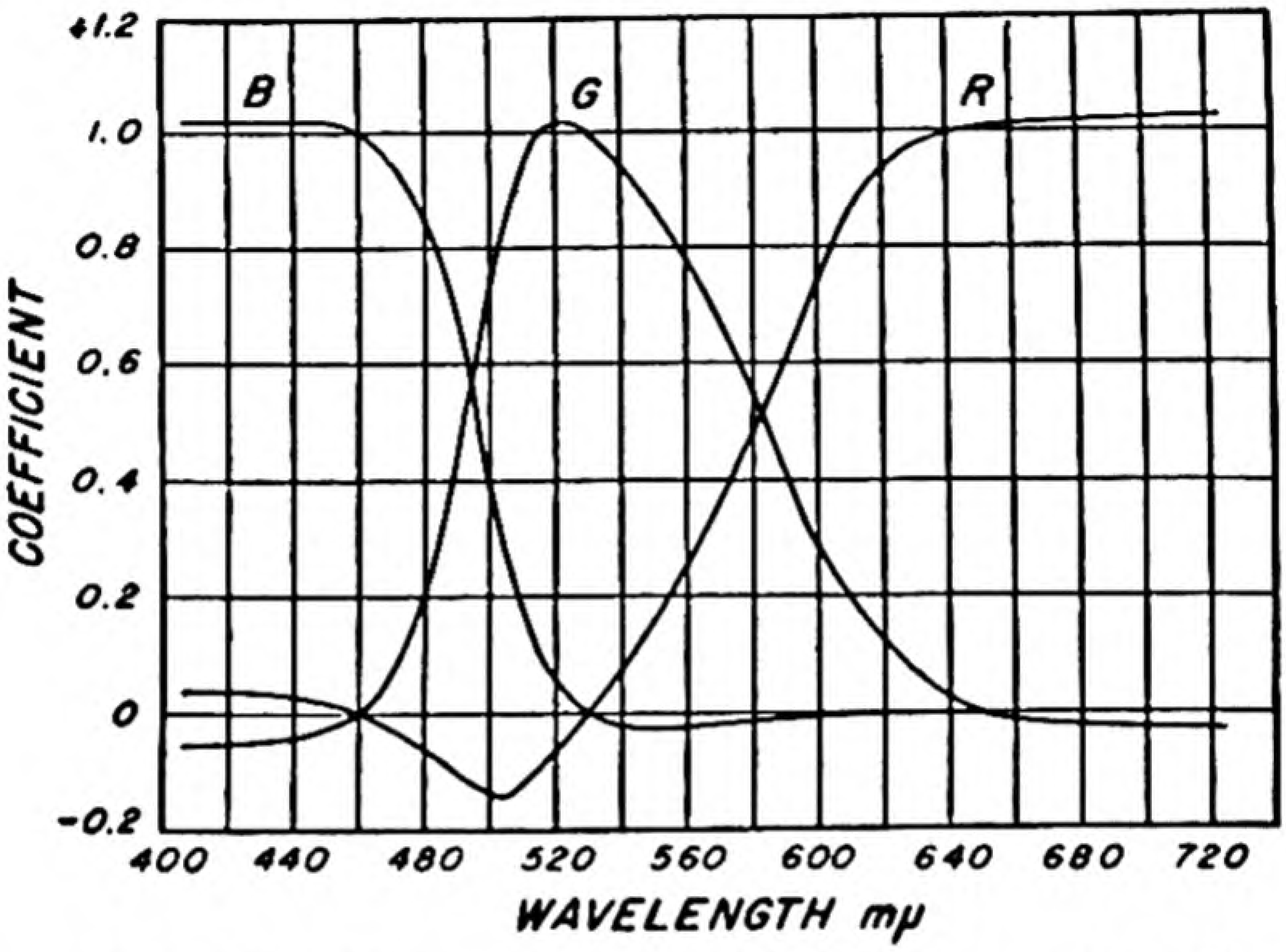


FIG. 35. Three-Component Mixture: The Proportions of the Three Primary Components *(R, G, B)* Required to Match the Spectral Colors. For a full explanation of the method of determining the coefficients, see text. (From L. T. Troland, *Principles of psychophysiology, Vol. II*, 1930, p. 160, by permission of D. Van Nostrand Company, Inc.)

Three-Component Mixture. Let us return to the law of intermediates. In our previous example we varied the hue of the mixture by varying the wave lengths of the components. It is also possible to vary the hue of the mixture, holding the wave lengths of the components constant and varying their relative intensities. Take yellow and green, for example. If the yellow light is intense and the green light is weak, the resultant mixture is predominantly yellow. Conversely, if the green light is relatively intense, the resulting mixture is predominantly green. By proper variation of the relative intensi-

ties, all the hues between yellow and green are obtained. By referring to Fig. 28, we can see that with two components we can produce *at most* the hues of half the color circle. If we add a third component, all the hues of the color circle can be produced by mixing these three components in their appropriate proportions. The selection of three such components is quite arbitrary, provided they are not all three selected from the same half of the color circle. A convenient triad, which has frequently been used in the investigation of stimulus mixture, consists of a red (650 mμ), a green (530 mμ), and a blue (460 mμ).

By the mixture of these three components in proper proportions, all the hues of the color circle can be obtained. In Fig. 35 are shown the proportions of the three components required to match all the spectral colors. A word should be said about the meaning of the units plotted. The components were first mixed in proportions to give white. Each of the three energies required for this mixture was taken as unity. Then, a coefficient of 0.5 indicates half the energy with which a component entered into the mixture for white, a coefficient of 0.7, 7/10 of the energy used in the mixture for white, and so on. Let us consider, for example, a spectral orange with a wave length of 600 mμ. To match this homogeneous light with the three components, red, green, and blue, the following fractions of the amounts which give white must be taken: about 0.7 R, about 0.3 G, and about 0.0 B. The negative values in Fig. 35 resulted from the attempt to match the spectral colors in saturation as well as in hue. It sometimes became necessary to add a fraction of one of the components to the homogeneous light being matched. This addition acted as a complementary to the spectral color and reduced its saturation. Thus, a green of 505 mμ could be matched by about 0.8 G + about 0.3 B, provided about 0.15 R was added to the homogeneous green of 505 mμ in order to reduce its saturation.

The remarkable fact that all the colors of the spectrum can be matched by mixtures of three components has served as the basis for one of the most widely accepted theories of color vision—the Young-Helmholtz theory. Because of the facts of stimulus mixture, it is unnecessary to assume that there are as many color receptors in the eye as there are spectral colors. In fact, according to the Young-Helmholtz theory, there are only three species of cones in

the retina. Each of these cones has its own visibility curve and produces its own particular color response. Just as the mixture of three lights gives all the colors, so the excitation of the three assumed types of cones accounts for all the chromatic colors. This theory accounts for more than the facts of stimulus mixture. Such well-known phenomena as color blindness and negative afterimages may be interpreted in terms of this theory.

AFTERIMAGES

Positive Afterimages. If we fixate a bright light and then quickly look at a dark wall, we ordinarily see an image of the object just fixated. Such an image which corresponds in its light-dark relations to the original stimulus is called a *positive afterimage.* It is ordinarily brief in duration, and its appearance depends on two main factors: the brightness of the stimulus and the brightness of the afterfield. A bright stimulus and a dark afterfield are conducive to the appearance of positive afterimages. Positive afterimages result from the fact that the response of the visual apparatus to a stimulus does not cease as soon as the stimulus is removed. If the afterfield is dark, there is little new stimulation to mask or interfere with the afterexcitation. That the positive afterimage is brief is fortunate; otherwise, as we turned our gaze from object to object we should have a continuous kaleidoscope of superimposed images. The important fact to remember is that this phenomenon constitutes evidence for inertia in the visual system.

Negative Afterimages. If we fixate a black figure on a white field for a period of time, say, about 30 seconds, and then turn to a homogeneous field, say, a gray or white wall, the stimulus figure will appear on the wall with the black-white relations reversed. Such an image is called a *negative afterimage.* These negative afterimages may be explained by changes in the sensitivity to light of different parts of the retina. During fixation of the dark figure, the retinal area on which it falls is less stimulated than the area which is stimulated by the white surface. During the period of fixation, therefore, the retinal areas stimulated by white become relatively less sensitive to light, while the area stimulated by the black figure becomes more sensitive to light. As a consequence, when the total retina is stimulated by a homogeneous surface, different parts of the retina re-

spond with different degrees of excitation. The area formerly stimulated by black responds at a higher level than the area formerly stimulated by white. Thus, there appears to be a light figure on a dark field.

Negative afterimages also occur after a chromatic stimulus has been fixated. Let us, for example, fixate a richly saturated blue pattern for 30 seconds and then turn to a well-illuminated white wall. An image of the pattern will appear, but its hue will be the complementary of blue, i.e., yellow. Such complementary afterimages may also be explained by adaptation in the framework of the Young-Helmholtz theory. Fixation of the blue pattern presumably led to maximum excitation and, hence, decreased sensitivity (p. 128) of the *B* cones. When the retina is stimulated by a homogeneous surface, that part of it formerly stimulated by the blue pattern now responds predominantly with the *R* and *G* cones, thereby producing yellow. Thus, the facts of complementary afterimages are in accord with the Young-Helmholtz theory.

Dark Adaptation

In the preceding section we have emphasized the importance of changes in the sensitivity of the retina to light: (1) increased sensitivity of the retina to light due to absence of stimulation in the dark is known as *dark adaptation;* (2) decreased sensitivity as a result of stimulation is called *light adaptation.*

The Dark-Adaptation Experiment. Increase in sensitivity to light as a result of time in the dark is usually measured under the following standard conditions. The subject is taken into a dark room and first required to "light adapt" by looking at a brightly illuminated surface for a period of time. By this procedure, the subject's sensitivity to light is greatly reduced, and a well-defined starting level is obtained from which the course of dark adaptation can be followed. After light adaptation, the subject's absolute threshold for a light stimulus of specific wave length and subtending a known angle on the retina is measured at frequent intervals. The minimum energy required for a visual response decreases as a function of time in the dark. Fig. 36 shows a family of dark-adaptation curves. On the abscissa, time in the dark is plotted; on the ordinate, the minimum brightness required for a visual response. Different curves are ob-

tained with different brightnesses of the light-adapting stimulus. Note that in all of the curves there is an initial steep drop in the threshold (increased sensitivity). Thereafter, the threshold continues to drop over a protracted period of time, though at a less rapid rate. Although most of the increases in sensitivity occur over the

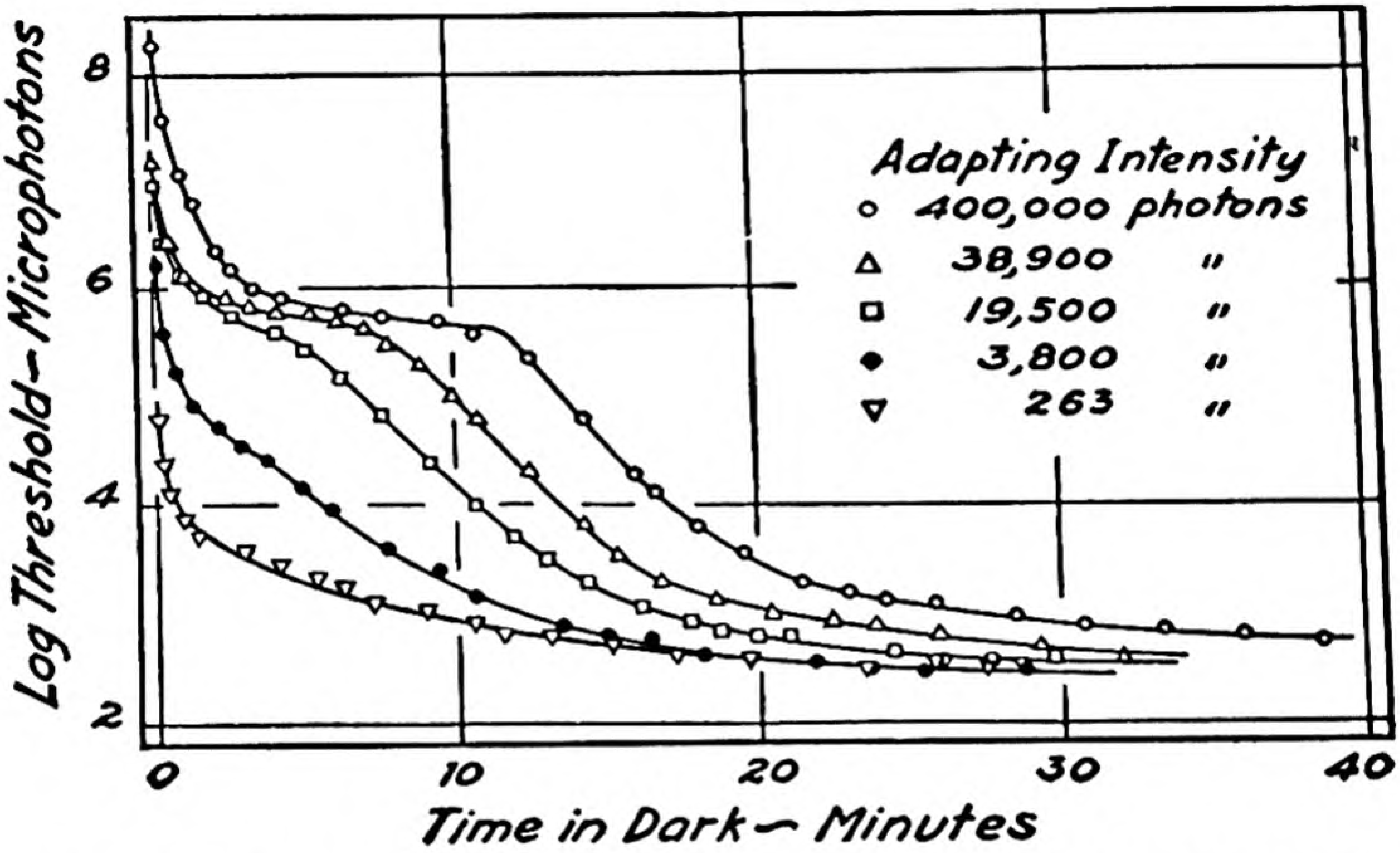


FIG. 36. A Family of Dark-Adaptation Curves. The intensity of the threshold stimulus is plotted against time in the dark. Note the upward shift of the curves with increase in the initial light-adapting intensity. (From S. Hecht, C. Haig, and A. M. Chase, The influence of light adaptation on subsequent dark adaptation of the eye, *J. Gen. Physiol.*, 1937, *20*:837, by permission of journal and authors.)

period of time pictured in the graph, some increase still occurs after several hours.

Dark Adaptation of Rods and Cones. Most of the curves in Fig. 36 consist of two branches: the first branch results from the steep initial drop in the threshold which levels off after about 5 minutes; the second branch also consists of a pronounced drop, less steep than the first, followed again by a level part which continues to drop slowly over a long period of time. The top function in Fig. 36 illustrates these two branches most clearly.

The two branches result from the difference in speed with which the cones and rods recover their sensitivity in the dark, and even more from the difference in the extent of dark adaptation for the two kinds of receptors. The first branch represents primarily cone adaptation, which is rapid but which levels off at a relatively high level of intensity. The second branch is due primarily to the adaptation of the rods, which is slower in onset, continues for a longer time, and represents a much larger change in sensitivity. Whereas the sensitivity of the cones is increased more than a hundredfold in the dark, the sensitivity of the rods is increased by a factor of more than 10,000.

The correlation of the two branches of the curve with the adaptation of cones and rods can be established by restricting the test stimulus either to the fovea (rod-free area) or to the periphery (virtually cone-free). If the stimulation is foveal, only the first of the two branches appears, whereas peripheral stimulation results in a curve corresponding to the second branch.

Further evidence for the interpretation of the total dark-adaptation curve in terms of two distinct processes comes from experiments in which stimulus lights of different wave lengths are used. As the visibility curves (Fig. 32) show, the rods are maximally sensitive to the shorter wave lengths, and the cones to the longer wave lengths. If a red homogeneous test patch is used to measure changes in sensitivity in the dark, the function closely resembles the upper branch of the total curve. On the other hand, if the test patch is violet, the lower branch of the curve is much more evident than the upper branch.

As we have seen, the light-sensitive substance in the rods—rhodopsin—is broken down by light. In the dark, this substance is regenerated. The increased concentration of rhodopsin accounts for the increase in sensitivity. This process of regeneration is slow, which explains the long period of time covered by the dark adaptation of the rods. The photochemical processes involved in the dark adaptation of the cones are still relatively obscure.

The fact that the rods are relatively insensitive to the reds at the extremes of the spectrum has an interesting application in dark adaptation. Suppose an individual wears goggles which act as a filter and let through only the very long wave lengths of light. Even

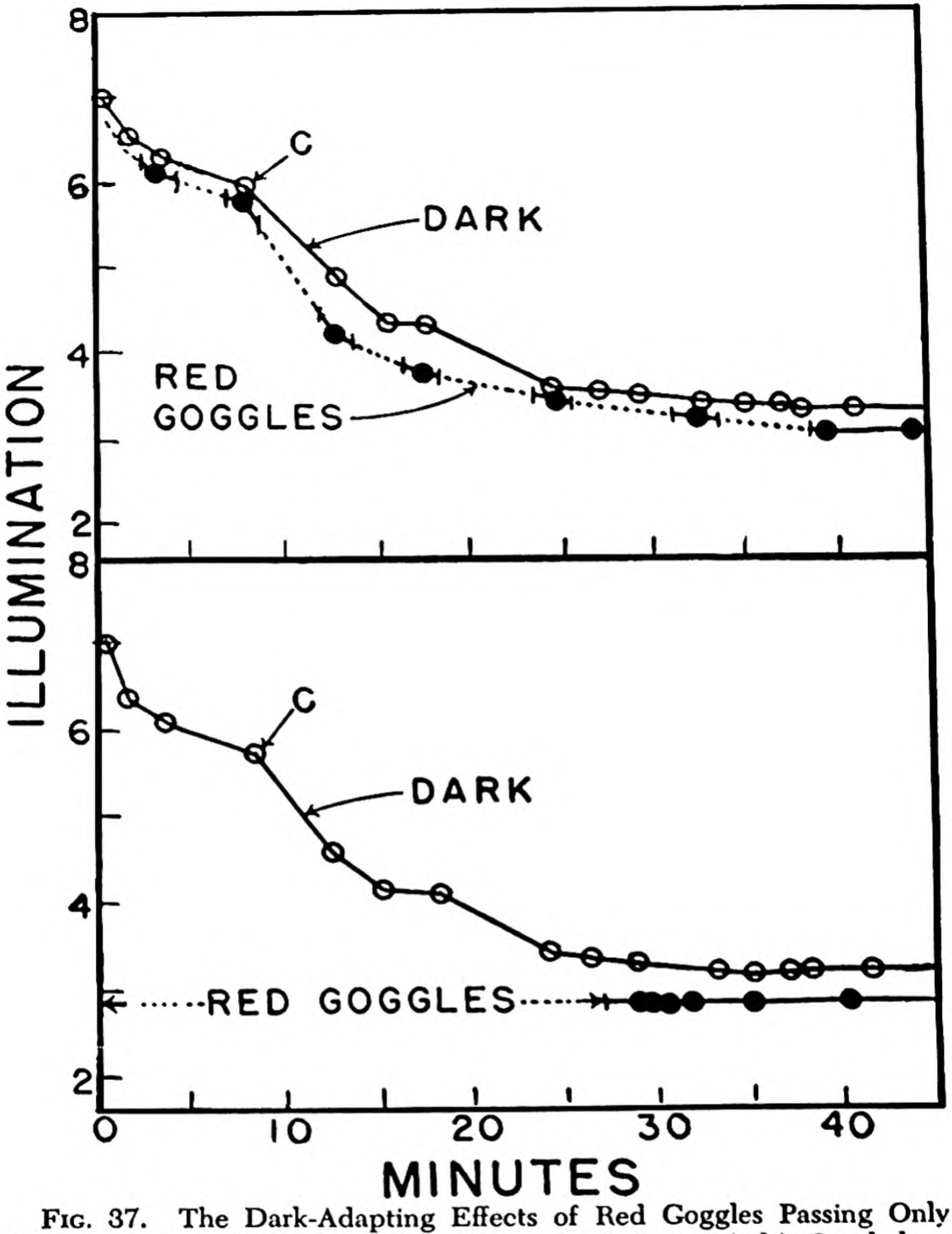


FIG. 37. The Dark-Adapting Effects of Red Goggles Passing Only Very Long Wave Lengths of Light. (From E. G. Boring (ed.), *Psychology for the armed services,* 1945, p. 54, by permission of Infantry Journal, Inc. After W. R. Miles, Red goggles for producing dark adaptation, *Fed. Proc.,* 1943, 2:112 f., by permission of the journal and author.)

if he is in a well-lighted environment, the rods are then free from stimulation and can, therefore, dark adapt. The red light, however, which stimulates the cones is sufficient for daylight vision, so that the individual can still use his eyes while "dark adapting." As in Fig. 37, the red goggles are extremely efficient in keeping out all light to which the rods are sensitive and allow rod adaptation as good as in complete darkness. Such goggles have been used to good advantage in night work. A dark-adapted individual on lookout duty, for example, may put goggles on and light a cigarette without disturbing his dark adaptation. Furthermore, he may wear them for a certain time before his duty begins and arrive at work adequately dark adapted.

In Fig. 37 the curves labeled *Dark* are dark-adaptation curves obtained under standard conditions. The letter *C* indicates the break between the cone and rod parts of the curve. In the upper graph the dotted line shows the dark adaptation of the rods by means of red goggles with continued cone vision. The tests were made in the dark (solid circles). The results in the lower diagram were obtained by having the subject wear the red goggles for 27 minutes and then testing him in the dark. The full effectiveness of the goggles is demonstrated by the fact that there were no further increases in sensitivity in the dark.

Determinants of the Rate and Amount of Dark Adaptation. There are many determinants of the specific forms which a dark-adaptation curve takes. Here is a list of some of the factors which need to be taken into account in planning a dark-adaptation experiment.

1. Intensity of light-adapting stimulus. The level to which the individual is light adapted importantly determines the initial value of the threshold taken in the dark, and, hence, the amount of time needed to reach complete adaptation. The family of curves in Fig. 36 illustrates this fact.
2. Retinal area stimulated. The retinal area stimulated is a critical factor. The relative number of rods and cones in the area stimulated by the test patch determines the relative prominence of the two branches of the total curve. In determining the dark-

adaptation curve as a function of retinal area, it is important that the subject maintain a steady point of fixation.

3. Size of test patch. This factor is clearly related to the retinal area stimulated. A large test patch cannot be confined to either a rod-free or a cone-free retinal area. A test patch larger than the foveal area results in a curve with both branches clearly present.
4. Wave length of the test patch. The difference in the visibility curves for rods and cones causes the shape of the dark-adaptation curve to vary with the wave length of the test patch. Long wave lengths result in a predominance of the cone branch of the curve, and short wave lengths in a predominance of the rod branch.

Light Adaptation

A period of time in the dark increases the sensitivity of the retina; a period of exposure to light decreases the sensitivity. To this decrease in sensitivity resulting from stimulation, the name *light adaptation* is given. The exposure of a dark-adapted retina to light results in a rapid elevation of the threshold. At first, the decrease in sensitivity is considerable and is exceedingly rapid. Thereafter, the threshold continues to rise but at a slower rate. If the adapting stimulus is intense, the eyes are maximally light adapted after a period of about 10 minutes and little further change in the threshold takes place. With a less intense adapting stimulus, the rate and amount of light adaptation are less.

A light-adaptation experiment may be performed under the following conditions. The subject is first fully dark adapted so that the process of light adaptation starts from a known level of sensitivity. After dark adaptation, the eye is exposed to a light of fixed intensity. (The rate and amount of light adaptation will, of course, depend on the intensity of the stimulus.) This exposure is continued for a brief interval, then the stimulus is turned off and the absolute threshold is determined as quickly as possible. This procedure is repeated for different durations of exposure to the light.

The process of light adaptation is difficult to chart. For one thing, it is so rapid that it is difficult to make a sufficient number of threshold determinations during the period in which its rate of change is greatest. Furthermore, there is a serious methodological

difficulty. When the stimulus to which the eye is being adapted is turned off so that a determination of the absolute threshold may be made, the reverse process of dark adaptation immediately sets in. The absolute threshold may thus be systematically too low.

Another way in which the visual apparatus adapts itself to variations in stimulus intensity is by changes in the size of the pupillary area. In general, the greater the brightness of a stimulus, the smaller is the pupillary area. The relationship between the logarithm of

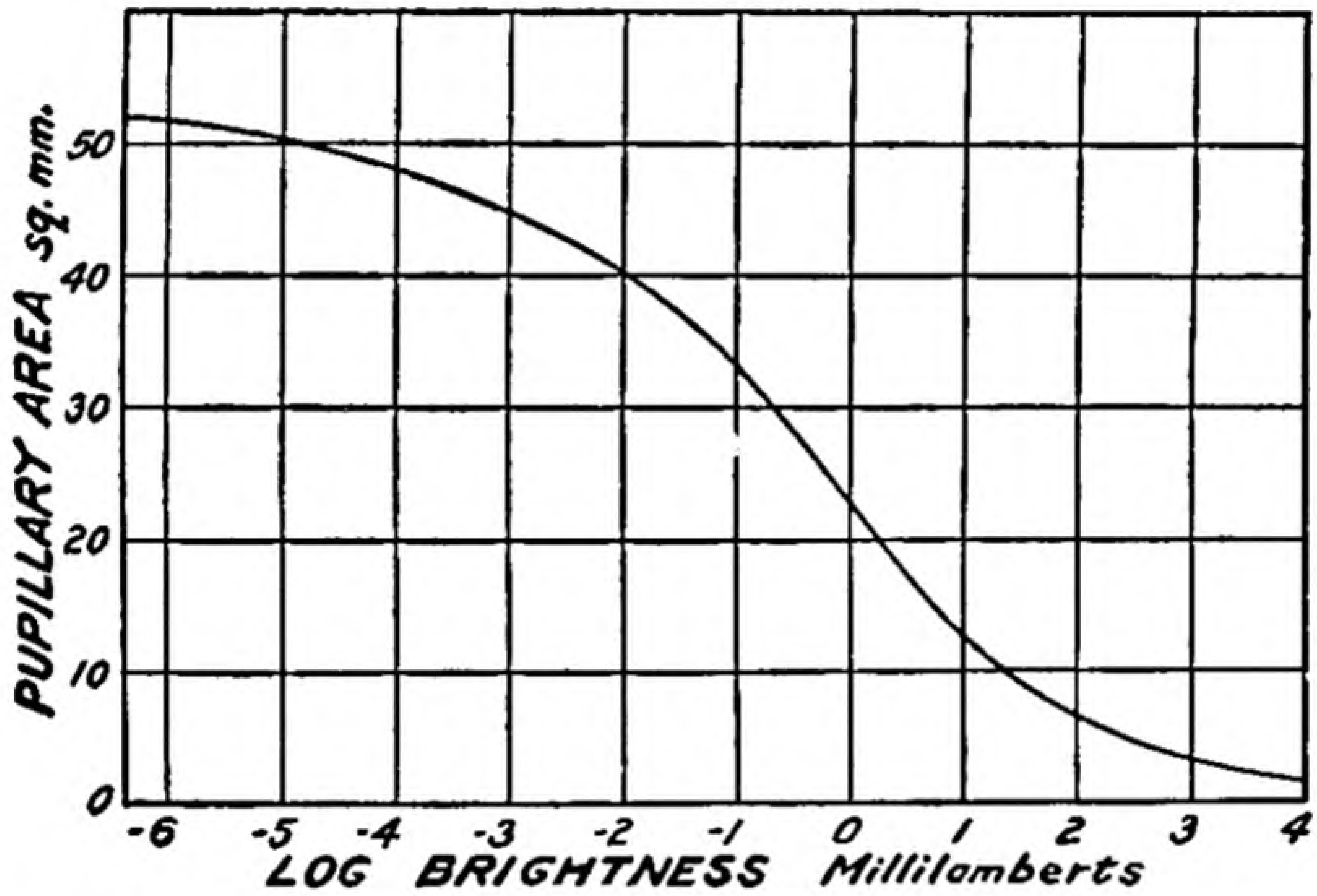


FIG. 38. The Relationship Between the Logarithm of the Brightness of the Visual Stimulus and the Size of the Pupillary Area. (From L. T. Troland, *Principles of psychophysiology, Vol. II,* 1930, p. 111, by permission of Van Nostrand Company, Inc.)

brightness as measured in millilamberts and the size of the pupillary area is depicted in Fig. 38. Over ten logarithmic units, the size of the pupillary area varies over a range of 50 square millimeters. Size of pupil is a factor which needs to be taken into account carefully in specifications of effective stimulus values in visual experiments. To hold the factor of pupil size constant, investigators have sometimes employed an artificial pupil smaller in size than the smallest "natural" size for the stimulus intensities used. In this manner, stimulus area can be held constant despite variations in brightness.

Visual Acuity

The Concept of Visual Acuity. One of the most important and useful characteristics of visual function is its spatial resolving power. By resolving power, we mean the capacity to discriminate different parts of the visual field from each other. This power of discrimination is very great and enables us to see very small objects and to respond to very small separations among objects in the visual field. *Visual acuity* is formally defined as the reciprocal of the visual angle subtended by the smallest object whose presence can be detected, or by the smallest separation between two objects which can be detected. The visual angle is measured in minutes of arc. This measure has the advantage of making the determination independent of the distance of the stimulus from the subject's eye. Many objects have been used to test visual acuity: single small objects, such as dots and hairlines; separations between two dots or two lines; minimum discriminable opening in a circle; the letter *C;* and a grating consisting of alternate light and dark bars.

No one value of maximal acuity (minimal angle) can be stated for the normal eye, for visual acuity is a function of the conditions under which the determinations are made. Under the best possible conditions, a hairline subtending an angle as small as 0.5 seconds has been detected. A more typical measure refers to the minimal detectable separation between two lines, a value of about 30 seconds, representing maximum acuity under favorable conditions. In medical eye examinations, an angle of 60 seconds defines normal acuity.

Acuity and Intensity. It is a commonplace of everyday experience that visual acuity is low in dim illumination. As night falls, even large objects and separations between objects can hardly be discriminated. The dependence of visual acuity on intensity of illumination has been carefully measured under controlled experimental conditions. Fig. 39 shows classic data depicting this relationship. The shape of the function is sigmoid. As intensity increases from an extremely low level, visual acuity at first increases slowly and then much more rapidly until rather high intensities are reached. With further increase in intensity, visual acuity increases very little.

There is reason to believe, from independent evidence, that the visual-acuity function, like the dark-adaptation function, should show two branches. In Fig. 39, the limb at the bottom of the curve

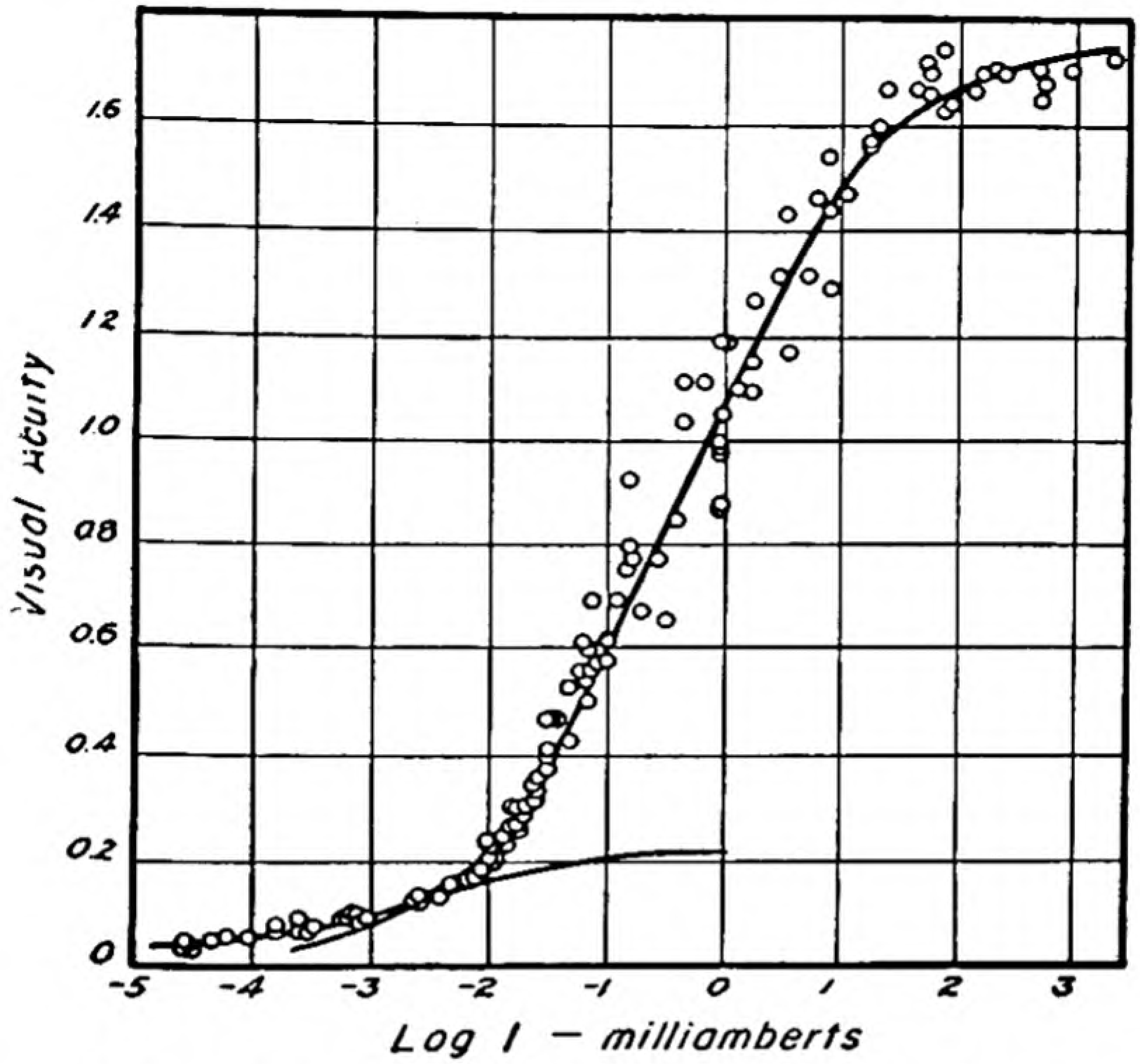


Fig. 39. The Relationship Between Visual Acuity and Intensity of Illumination. The limb at the bottom of the curve indicates the course of acuity for the rods. The main branch of the curve is determined primarily by the response of the cones. (From S. Hecht, Vision: II. The nature of the photoreceptor process. In C. Murchison (ed.), *A handbook of general experimental psychology*, 1934, p. 774, by permission of Clark University Press.)

indicates the course of acuity for the rods. Once the cones go into action, their response principally determines the acuity.

We must not conclude that acuity always increases with intensity. The brightness relation between the test object and its surround is important. If the surround is dark and the test objects are two bars illuminated, increase in the brightness of these bars improves acuity only over a limited range. At the higher intensities,

acuity actually goes down. Since under such conditions most of the retina is undergoing dark adaptation, a glare effect is produced. The glare probably results from the fact that there is no stable level of adaptation for the retina as a whole, with most of the retina weakly illuminated and the two small test objects having considerable brightness. Except under such circumstances, increase in illumination does progressively improve acuity.

Fundamentally, visual acuity may be regarded as a form of brightness discrimination. When, for example, two dark bars are separated by a light space, the detection of this separation depends on the difference between the brightness of the bars and the brightness of the interspace. The seeing of the two bars as separate is, after all, the response to the brightness difference. If we made the dark test bars lighter and lighter without altering the brightness of the interspace, the detection of their separateness would become more and more difficult.

Now, suppose we increase the illumination on both the dark test bars and the light interspace. The ratio of the brightness of the test bars to the brightness of the interspace remains the same, and yet visual acuity increases. This increase in acuity is due to the fact that brightness discrimination is better at the higher level of illumination.

Acuity and Retinal Area. Let us look at visual acuity from a somewhat different point of view. The retina consists of a mosaic of discrete receptor elements. These receptor elements are connected to nerve cells. In most cases, several receptors are connected to one nerve cell, and thus form a single functional unit. In other cases, only a *single* receptor initiates impulses in a nerve cell. The mosaic of the retina refers to these functional elements rather than just to the spatial arrangement of the receptors.

Visual acuity is a measure of the grain of this mosaic. If visual acuity is good, the grain must be fine; if visual acuity is poor, the grain must be coarse. If the retina varies in grain from one area to another, it follows that visual acuity should depend on the particular retinal area stimulated. This is, indeed, the case.

In the fovea, which is tightly packed with cones, each of which has its own nerve connection, the grain of the retinal mosaic is exceedingly fine. In the periphery, where rods with shared nerve connections predominate, the retinal mosaic is coarse. In close

correspondence with this neurological picture, visual acuity is best in the fovea and becomes increasingly poor as we go out to the periphery. The superior visual acuity of the fovea is further aided by the fact that optically the clearest visual image is cast in the center of the retina.

Spatial and Temporal Summation

Spatial Summation. As already indicated, the retina is a mosaic with a complicated network of interconnections. As a result, the effects of the stimulation of a specific spot on the retina are not confined to that spot alone. For example, as a very small light stimulus is increased in size without increasing its intensity, its brilliance increases. Brilliance increases because increase in the size of the stimulus brings in more and more receptor elements whose effects summate. Furthermore, the threshold of a spot closely adjacent to one stimulated is lowered.

Demonstrable summation effects are restricted to very small areas on the retina. In the fovea they extend over less than a degree of visual angle. In the periphery they may extend over as much as two degrees. This greater range of spatial summation in the periphery is another example of the fact that the mosaic of the periphery is coarser in grain than that of the fovea.

Temporal Summation: Roscoe-Bunsen Law. Retinal summation effects are not only spatial but also occur over short periods of time. Over a short duration range, increased time of action of the stimulus may compensate for decreased intensity, so that over this short range of durations a constant perceptual effect may be maintained if $It = C$, where I stands for intensity, t for time, and C for a constant. The temporal range over which this relationship holds varies with the constant level of brilliance which is obtained by combinations of I and t. If a threshold effect or a very low brilliance is required, the effective temporal range may be as long as 1/10 seconds. When a high constant effect is demanded of the stimulus, the temporal range is shorter, about 2/100 seconds. The generalization that $It = C$ is known as the *Roscoe-Bunsen law.* A breakdown of the Roscoe-Bunsen law is implied in the fact of the absolute threshold. If the law held over an indefinite range, we

could see a light of any dimness provided it were exposed for a sufficiently long time.

Talbot's Law. An intermittent stimulus may be seen as continuous. This effect may be demonstrated by interrupting a beam of light with a rotating disk. This disk has segments cut out of it so that part of the time the light may pass through. Let us assume that the light is allowed to pass only 50 percent of the time. If the disk is rotated slowly so that the light is interrupted only two or three times a second, the subject sees the light as interrupted, i.e., he sees alternations of light and dark. As the speed of rotation of the disk, and hence the rate of interruption, increases for a time, the subject still sees the light as interrupted. With further increases in the speed of rotation, fluctuations in brilliance become less and less marked, and finally a point is reached at which the light appears as continuous. The brilliance of this continuous light will be the same as if the total amount of light had been uniformly distributed over a whole revolution of the disk. In other words, the brilliance of the rotating disk, passing the light 50 percent of the time, would match a steady light whose intensity is half that of a single flash. In general, if a light is interrupted frequently enough to result in fusion, the following generalization about its brilliance holds. If the light flashes are on *p* percent of the time, the brilliance of the interrupted stimulus is the same as that of a continuous stimulus whose intensity is *p* percent of the intensity of a single flash. This generalization is known as *Talbot's law* and has received frequent experimental confirmation.

The Critical Fusion Frequency and Its Determinants. As we have seen, there is a certain minimum frequency of interruption necessary for an interrupted stimulus to be seen as continuous. This minimum frequency of interruption at and above which the interrupted light appears as steady is known as the *critical fusion frequency* (c.f.f.). Below this frequency, the interrupted light is seen to flicker. There is no single value for the c.f.f.; rather it depends on a number of variables, of which intensity of the flash is among the most important.

Fig. 40 shows c.f.f. as a function of the logarithm of intensity, measured under conditions where light and dark periods of a cycle were equal. Let us first consider the curve defined by the open

circles. The relationship between c.f.f. and intensity is sigmoid. Over a considerable range, however, relation between c.f.f. and log *I* is nearly linear, i.e., c.f.f. is proportional to log *I*. The logarithmic relationship between c.f.f. and intensity is known as the *Ferry-Porter law.*

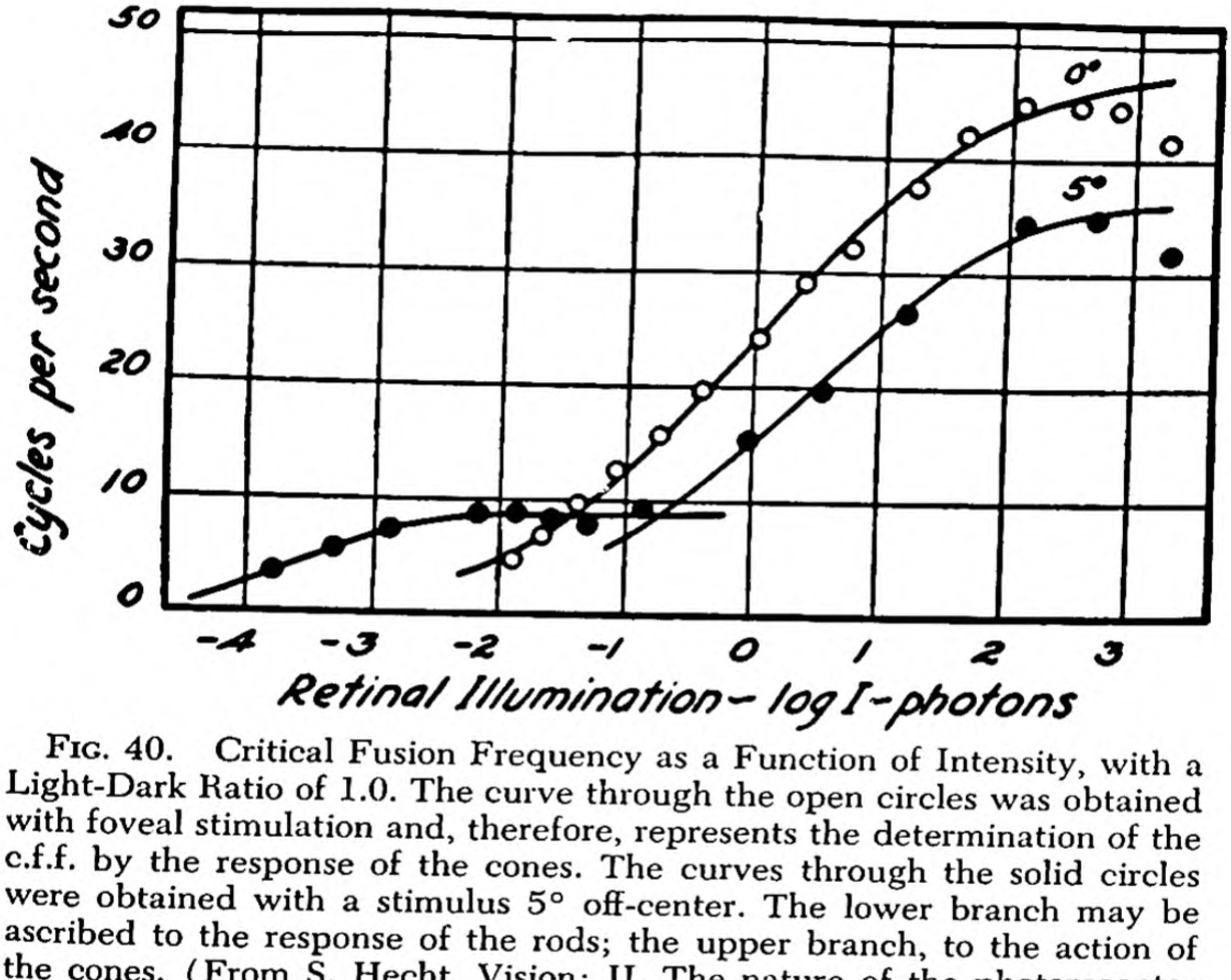


FIG. 40. Critical Fusion Frequency as a Function of Intensity, with a Light-Dark Ratio of 1.0. The curve through the open circles was obtained with foveal stimulation and, therefore, represents the determination of the c.f.f. by the response of the cones. The curves through the solid circles were obtained with a stimulus 5° off-center. The lower branch may be ascribed to the response of the rods; the upper branch, to the action of the cones. (From S. Hecht, Vision: II. The nature of the photoreceptor process. In C. Murchison (ed.), *A handbook of general experimental psychology,* 1934, p. 750, by permission of Clark University Press.)

The curve through the open circles in Fig. 40 was obtained by confining the stimulus to the fovea. It represents, therefore, the dependence of c.f.f. on the action of the cones. If c.f.f. is determined for a retinal area outside the fovea, a function with two limbs is obtained. The two curves passing through the filled circles in Fig. 40 were obtained with a stimulus at 5° off-center. The low branch obtained with weak intensities may be ascribed to the action of the

rods; the upper branch of the curve represents primarily the action of the cones. Thus, the c.f.f. as a function of intensity takes a different course for the two types of receptors.

The results in Fig. 40 were obtained by interrupting the light half the time. 50 percent light and 50 percent dark (or a light-dark

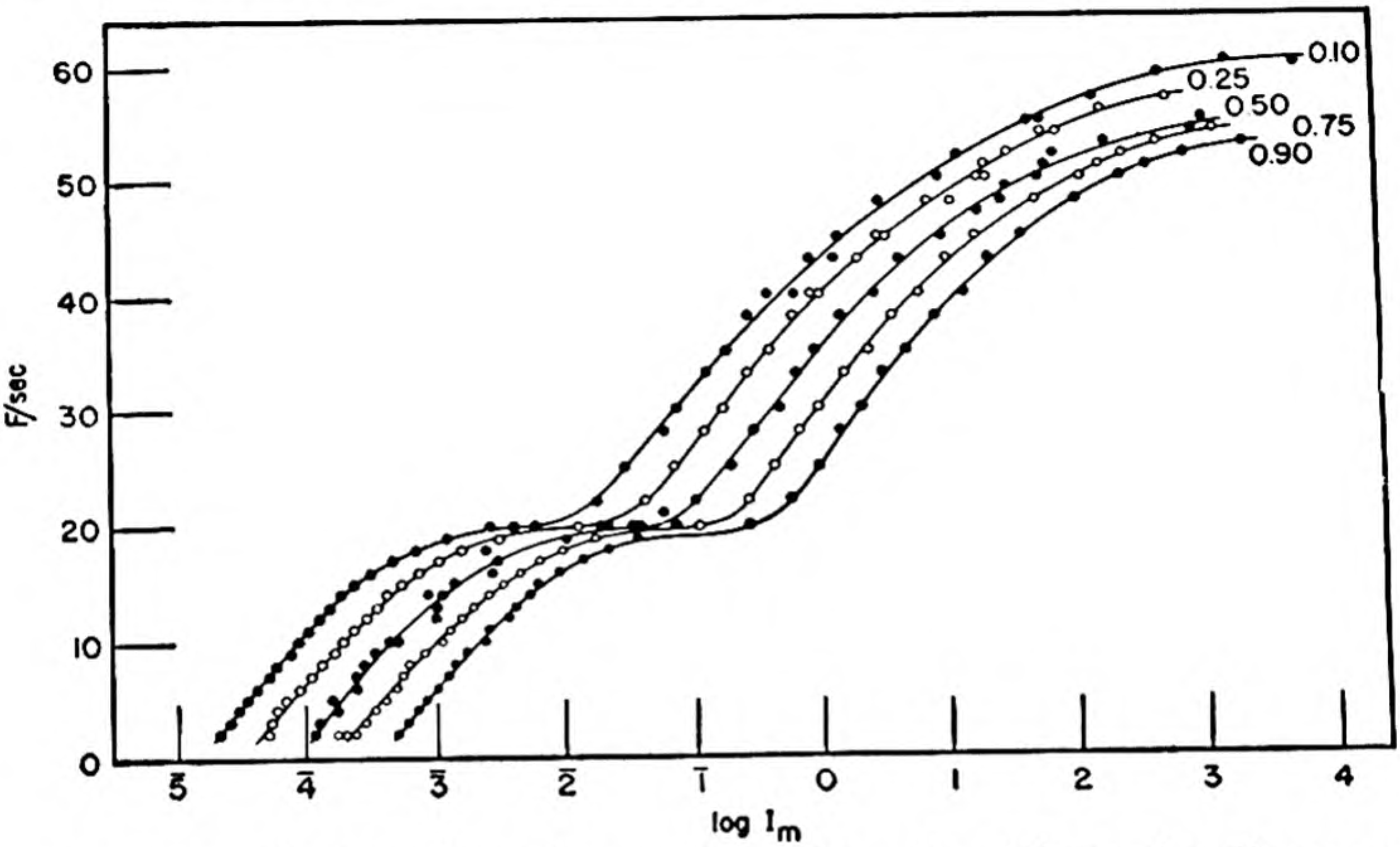


FIG. 41. This Family of Curves Shows How the Relationship Between Critical Fusion Frequency and Intensity Varies with the Light-Dark Ratio in the Total Cycle. (From W. J. Crozier and E. Wolf, Theory and measurement of visual mechanisms: V. Flash duration and critical intensity for response to flicker, *J. Gen. Physiol.*, 1941, *24*:639, by permission of journal and authors.)

ratio of 1.0) represent, of course, only one of many possible combinations of light and dark periods. Much lower and much higher percentages of light in the total cycle have been used. The relationship between c.f.f. and the logarithm of intensity varies with the percent of time that the light is on in the total cycle. Fig. 41 shows this fact. In general, the greater the percentage of time that the light is on, the lower is the c.f.f. The c.f.f. of a light of a given intensity which is on 90 percent of the time is lower than that of a light of the same intensity which is on only 50 or 10 percent of the time.

The Duplicity Theory

Throughout this chapter we have seen again and again the differences between the responses of the two types of receptors in the retina—the rods and the cones. When we have studied various visual functions, such as visibility, light and dark adaptation, visual acuity, flicker, and color vision, we have always encountered facts reflecting their difference in function. Other evidence, coming, for example, from anatomy, physiology and pathology, fully supports these experimental findings. The statement of the duplicity theory—that there are two functional types of receptors in the retina—is now an established fact.

Special Problems of Control in Visual Experiments

The visual apparatus is extremely sensitive and responsive to even the slightest variations in external and internal conditions. Visual experimentation requires most careful control of a variety of factors of which some of the most important will be listed below. Let us follow the path of the light stimulus from its source to the retinal image and note on the way what some of the problems of experimental control are.

Intensity of the Stimulus. The accuracy with which the intensity of the visual stimulus needs to be specified depends, of course, on the particular problem. The investigation of the absolute threshold or of the relation of c.f.f. and intensity, of course, requires explicit and careful specifications of the intensity of the stimulus. On the other hand, verification of the fact of complementary colors does not require exact photometric measurements of the illumination on the color disks. Whenever specifications of intensity are given, the experimenter must carefully distinguish between radiometric and photometric measures.

Spectral Composition of the Stimulus. The remarks made above about the importance of exact measurements in different experimental situations apply here too. In determining visibility curves, the necessity for exact specification of wave lengths is obvious. In experiments involving the use of visual patterns illuminated by white light, the spectral composition of that white light need not be specified. Specification frequently is implied by naming the

particular source of illumination, such as the brand of light bulb. Where wave length must be controlled, filters are frequently used. A typical filter consists of two glass plates enclosing a gelatinous substance which absorbs certain wave lengths while transmitting others. Filters vary in degree of selectivity, and for precise work it is necessary to have available their transmission characteristics. Typically, a curve showing degree of transmission as a function of wave length is plotted.

Surround of a Test Patch. There are many experiments in which the surround of the test stimulus or test patch is important. This is particularly true when the test patch is small. We recall the special effects (glare) encountered in the determination of visual acuity when a brightly illuminated small object appears in a dark surround. A phenomenon such as glare is, of course, in itself an interesting problem. Even if the effect of the surround is not itself the focus of investigation, it should be taken carefully into account.

Size of Pupillary Area. The size of the pupillary area is important for two principal reasons. First of all, the size of the pupil helps to determine the amount of illumination which falls upon the retina. We recall that it is for this reason that a special unit of illumination—the photon—was devised. Furthermore, the size of the pupil has an effect on the sharpness of the retinal image. Down to a certain point, as the area of the pupil is decreased, the definition of the retinal image is sharpened. However, as the pupillary area decreases further, diffraction effects enter and serve to blur the image. For these reasons, in experiments in which the amount of illumination or the sharpness of the image (visual acuity) is important, pupillary area must either be specified or controlled by an artificial pupil.

Retinal Area Stimulated. The facts of the duplicity theory underscore the importance of controlling and exactly specifying the retinal area stimulated by the test patch. In order to stimulate a particular area of the retina, it is, of course, necessary to control the size and distance of the test patch. The area of the retinal image corresponding to a given object is inversely proportional to the square of the distance of the object from the eye. Applying this relationship, it should be remembered that for foveal stimulation the stimulus should not cover an area larger than about 2° of visual angle. If we are to be sure that a test object falls on a particular

part of the retina, we must minimize the movements of the eye. This is achieved first by holding the head in a chin rest or similar device and then requiring the subject to maintain steady fixation on a point provided for that purpose.

State of Adaptation of the Retina. In many experiments visual functions are determined after the eye has been adapted to the particular viewing situation. In an experiment on visual acuity, for example, it is assumed that the retina of the subject has come to a steady level of adaptation before the measurements are made. Thus, by specifying the level of illumination, we are adequately controlling the degree of adaptation. In some experiments, however, adaptation is itself the focus of interest. In such experiments, adaptation is explicitly measured by determining the absolute threshold.

Experiment VI
Stimulus Mixture

Purpose. To verify the two principal laws of stimulus mixture—the law of intermediates and the law of complementaries.

Materials. A set of color disks and one or more color wheels are used. Color disks can be commercially obtained or may be cut from a set of colored papers. For the purposes of this experiment, it is not necessary to know the photometric values of these stimuli. Furthermore, it is important to realize that such colored papers do not reflect homogeneous wave lengths.

A color wheel is a motor-driven metal disk, so constructed that the colored papers may be attached to it. The speed of the motor may usually be adjusted by means of a rheostat.

Procedure: The Law of Intermediates. This law states: The hue of the mixture of any two homogeneous components is intermediate on the color circle between the hues of those components. The mixture is less saturated than one or both of the components. The hue of the mixture is closer to that component which has the greater intensity.

To verify this law, choose first a red and yellow color disk. Interleave them on the color wheel, with only a small proportion of the yellow color showing. This means that the intensity of the red light will be greater than that of the yellow light. Adjust the speed of the motor well above the critical fusion frequency and observe the hue of the mixture.

Next, change the proportion of the red and yellow to fifty-fifty and again cbserve the mixture. How has it changed in hue and saturation? Finally, use only a small fraction of red and make the same observations.

If two color wheels are available, place a red and yellow disk on one of them as before, and a well saturated orange *plus* a white disk on the other wheel. Find the proportions of red and yellow required to match the *hue* of the orange disk. Then, try to obtain a *saturation* match. To obtain a saturation match, it will be necessary to vary the proportion of white added to the orange until a satisfactory match to the saturation of the red-yellow mixture is obtained.

Repeat this procedure with the following pairs of disks: yellow and green, green and blue.

Treatment of Results. For every mixture, the proportions of disks used should be carefully measured and recorded. The proportions can be readily determined by means of a protractor disk which is calibrated either in percentages or degrees. The results of each mixture should be recorded in the form of a simple "color equation," e.g.,

$$X\%\ \textit{Red} + (100 - X)\%\ \textit{Yellow} = Y\%\ \textit{Orange} + (100 - Y)\%\ \textit{White}$$

We may note that if the components were homogeneous lights, such an equation would fully specify the properties of one color in terms of the wave lengths and intensities of two others.

Procedure: The Law of Complementaries. This law states: If two hues are diametrically opposite each other on the color circle, the mixture of their stimuli in the proper intensity ratio appears gray.

The law may be verified by means of the following pairs of stimuli:

Yellow and blue
Green-blue and orange
Red and blue-green

The two disks of one of these pairs of stimuli are interleaved on the color wheel. Also interleaved on the same color wheel are a small black and a small white disk. The proportions of the two chromatic disks are adjusted until a gray mixture results. The proportions of the black and white disks are then adjusted to obtain a brilliance match. (Although it is preferable to mount the disks in this way, it may be more convenient to mount the black and white disks on a separate wheel.)

We must remember that the color disks do not reflect homogeneous light. For this reason, it may not be possible to obtain gray by the mixture of only two "components." We may still obtain a gray, however, by adding a third chromatic disk complementary to the mixture of the first two. Thus, the mixture of red and blue-green closest to gray may still appear tinged with blue. By adding a small amount of yellow, which is complementary to blue, a gray mixture may finally be produced.

The student should verify that he can readily obtain a gray with three components if they are not all three selected from the same half of the color circle.

Treatment of Results. The mixtures should again be specified in terms of "color equations." On one side should appear the proportions of chromatic disks used to obtain gray, on the other side the proportions of black and white giving a brilliance match. Thus,

$$X\%\ \mathit{Yellow} + (100 - X)\%\ \mathit{Blue} = Y\%\ \mathit{Black} + (100 - Y)\%\ \mathit{White}$$

Experiment VII

Visual Acuity[4]

Purpose. To measure two types of visual acuity: (1) *minimum visible acuity,* i.e., the smallest single object of a given kind that can just be detected; (2) *minimum separable acuity,* i.e., the minimum separation between two objects of the same kind that can be detected.

Minimum Visible Acuity

Materials. A white circular card, about 4 inches in diameter, is ruled with a single black line (india ink). The width of this line should be 1/64 of an inch and its length, 1 inch. (This width is somewhat arbitrarily chosen, and a narrower line would be preferable. The important point is that the width of the line should be exactly measured.) A yardstick for measuring distances and a lamp for varying the illumination of the card are required.

Procedure. In a room with ordinary illumination, the experimenter holds the stimulus card in a fixed position, with the black line vertical. The subject is initially close to the experimenter but he backs away until he can just see the black line. The experimenter then requests the subject to close his eyes and quickly rotates the card to a new position. He then asks the subject to indicate the inclination of the just visible line to the new position. This procedure serves as a check on the subject's report. For the experimenter's convenience, a line is also drawn on the back of the card in the same position as in the front. The distance between the stimulus card and the subject's eye is then measured to the nearest inch. It may be advisable to use both an approaching and receding (ascending and descending) series. At least ten such determinations should be made. The larger the number of determinations, the more reliable will be the resulting measure of visual acuity.

The whole procedure should be carried out at two levels of illumination. The first has already been specified, viz., ordinary room illumina-

[4] We are indebted to Professor K. U. Smith for suggesting this experiment.

tion. A second level may be obtained by casting the light of a desk lamp on the stimulus card.

Treatment of Results. In order to compute the visual acuity, it is necessary to express the width of the stimulus line and the mean distance at which it is just seen in the same units of length. When they are so expressed, the ratio between the width of the line and the mean distance equals the tangent of the minimum visible angle subtended by the object. From a table of trigonometric functions, the angle corresponding to the obtained tangent can be read. The value of this angle should be expressed in minutes, and its reciprocal computed. This reciprocal of the angle in minutes is the desired value of visual acuity. The values of visual acuity thus obtained should be compared for the two levels of illumination.

Minimum Visible Separation

Materials. A circular card 4 inches in diameter is used as before, but this time it is ruled with two parallel black lines so that there is a white interspace of 1/32 inch between them. The lines themselves should be 1/32 inch in width and 1 inch long. Yardstick and lamp are used as before.

Procedure. The same procedure is used as above, except the subject must now report when he can just see the two lines as separate. (Note that the control procedure of rotating the card is not applicable in this part of the experiment.) Again, several determinations are made at each of the two levels of illumination.

Treatment of Results. The minimum separable acuity is computed in much the same way as the minimum visible acuity. The width of the interspace between the two black lines is divided by the mean distance at which the two lines are just seen as separate. The reciprocal of the visual angle expressed in minutes is again the measure of acuity. The visual acuity at the two levels of illumination should again be compared.

General References

Bartley, S. H., *Vision: a study of its basis,* New York: D. Van Nostrand Co., 1941.

Bartley, S. H., Studying vision. In T. G. Andrews (ed.), *Methods of Psychology,* New York: John Wiley and Sons, 1948.

Boring, E. G., *Sensation and perception in the history of experimental*

psychology, New York: D. Appleton-Century Co., 1942, Chaps. 4 and 5.

Graham, C. H., Vision: III. Some neural correlations. In C. Murchison (ed.), *A handbook of general experimental psychology,* Worcester: Clark University Press, 1934.

Hecht, S., Vision: II. The nature of the photoreceptor process. In C. Murchison (ed.), *A handbook of general experimental psychology,* Worcester: Clark University Press, 1934.

Morgan, C. T., *Physiological psychology,* New York: McGraw-Hill Book Co., 1943, Chaps. 9 and 10.

Troland, L. T., *Principles of psychophysiology,* Vol. II, New York: D. Van Nostrand Co., 1930.

Troland, L. T., Vision: I. Visual phenomena and their stimulus correlations. In C. Murchison (ed.), *A handbook of general experimental psychology,* Worcester: Clark University Press, 1934.

Visual Stimulus

Hardy, A. C., and Perrin, F. H., *The principles of optics,* New York: McGraw-Hill Book Co., 1932.

Illuminating engineering nomenclature and photometric standards, Illuminating Engineering Society, New York, 1947.

Southall, J. P. C., *Introduction to physiological optics,* New York: Oxford University Press, 1937.

Dimensions of Color

Boring, E. G., The psychophysics of color tolerance, *Amer. J. Psychol.,* 1939, *52*:384-394.

Boring, E. G., Langfeld, S. H., and Weld, H., *Foundations of psychology,* New York: John Wiley and Sons, 1948, Chap. 12.

Crozier, W. J., On the law for minimal discrimination of intensities: IV. $\triangle I$ as a function of intensity, *Proc. Nat. Acad. Sci.,* 1940, *26*:382-389.

Hecht, S., Peskin, J. C., and Patt, M., Intensity discrimination in the human eye: II. The relation between $\triangle I/I$ and intensity for different parts of the spectrum, *J. Gen. Physiol.,* 1938, *22*:7-19.

Hecht, S., Shlaer, S. S., and Pirenne, M. H., Energy, quanta, and vision, *J. Gen. Physiol.,* 1942, *25*:819-840.

Hecht, S., and Williams, R. E., The visibility of monochromatic radiation and the absorption spectrum of visual purple, *J. Gen. Physiol.,* 1922, *5*:1-34.

Purdy, D. M., The Bezold-Brücke phenomenon and contours for constant hue, *Amer. J. Psychol.*, 1937, *49*:313-315.

Stimulus Mixture

Priest, I. G., Note on the relations between the frequencies of complementary hues, *J. Opt. Soc. Amer.*, 1920, *4*:402-404.

Sinden, R. H., Studies based on spectral complementaries, *J. Opt. Soc. Amer.*, 1923, *7*:1123-1154.

Afterimages

Berry, W., Color sequences in the after-images of white light, *Amer. J. Psychol.*, 1927, *38*:584-596.

Dark Adaptation

Crozier, W. J., The theory of the visual threshold: II. On the kinetics of adaptation, *Proc. Nat. Acad. Sci.*, 1940, *26*:334-339.

Hecht, S., Haig, C., and Chase, A. M., The influence of light adaptation on subsequent dark adaptation of the eye, *J. Gen. Physiol.*, 1937, *20*:831-850.

Mandelbaum, J., and Mintz, E. U., The sensitivities of the color receptors as measured by dark adaptation, *Amer. J. Ophthal.*, 1941, *24*:1241-1254.

Wald, G., and Clark, A. B., Visual adaptation and chemistry of the rods, *J. Gen. Physiol.*, 1937, *21*:93-105.

Light Adaptation

Crozier, W. J., The theory of the visual threshold: I. Time and intensity, *Proc. Nat. Acad. Sci.*, 1939, *26*:54-60.

Wright, W. D., The measurement and analysis of colour adaptation phenomena, *R. Soc. Lond. Proc.*, 1934, *115B*:49-87.

Visual Acuity

Fry, G. A., and Cobb, P., A new method for determining the blurredness of the retinal image, *Trans. Acad. Ophthal., Otolaryng.*, 1935.

Hecht, S., and Mintz, E. U., The visibility of single lines at various illuminations and the retinal basis of visual resolution, *J. Gen. Physiol.*, 1939, *22*:593-612.

Luckiesh, M., and Moss, F. K., *The science of seeing*, New York: D. Van Nostrand Co., 1938.

Wilcox, W. W., The basis of the dependence of visual acuity on illumination, *Proc. Nat. Acad. Sci.*, 1932, *18*:47-56.

Spatial and Temporal Summation

Adrian, E. D., and Matthews, R., The action of light on the eye: II. The processes involved in retinal excitation. III. The interaction of retinal neurons, *J. Physiol.*, 1928, *64*:279-301; 1928, *65*:273-398.

Bartley, S. H., Subjective flicker rate with relation to critical flicker frequency, *J. Exper. Psychol.*, 1938, *22*:388-394.

Bartley, S. H., Some factors in brightness discrimination, *Psychol. Rev.*, 1939, *46*:337-358.

Beitel, R. J., Spatial summation of subliminal stimuli in the retina of the human eye, *J. Gen. Psychol.*, 1934, *10*:311-327.

Crozier, W. J., and Wolf, E., Theory and measurement of visual mechanisms: IV. Critical intensities for visual flicker, monocular and binocular, *J. Gen. Physiol.*, 1941, *24*:505-534.

Crozier, W. J., and Wolf, E., Theory and measurement of visual mechanisms: V. Flash duration and critical intensity for response to flicker, *J. Gen. Physiol.*, 1941, *24*:635-654.

Crozier, W. J., and Wolf, E., Theory and measurement of visual mechanisms: VI. Wave-length and flash duration in flicker, *J. Gen. Physiol.*, 1941, *25*:89-110.

Crozier, W. J., and Wolf, E., Theory and measurement of visual mechanisms: VII. The flicker response function outside the fovea, *J. Gen. Physiol.*, 1941, *25*:369-379.

Granit, R., Comparative studies on the peripheral and central retina: I. On interaction between distant areas in the human eye, *Amer. J. Physiol.*, 1930, *94*:41-50.

Granit, R., and Davis, W. A., Comparative studies on the peripheral and central retina: IV. Temporal summation of subliminal visual stimuli and the time course of the excitatory after-effect, *Amer. J. Physiol.*, 1931, *98*:644-653.

Duplicity Theory

Crafts, L. W., Schneirla, T. C., Robinson, E. E., and Gilbert, R. W., *Recent experiments in psychology*, New York: McGraw-Hill Book Co., 1938, Chap. 9.

»»» 7 «««

PERCEPTION OF COLOR

THE phenomenal world is full of colors. We are surrounded by an ever changing kaleidoscope of brilliances and hues. However great the variety of our color experiences, it is small as compared with the constant changes in the stimulation of our visual receptors. Every time we turn our head, approach or recede from an object, the size, intensity, and pattern of the retinal image change, often radically. Yet, the perceptual world in which we move and to which we respond has great stability and, in spite of rich variety, much constancy. Out of the rich flux of sensory stimulation, the organism constructs, as it were, a stable and orderly perceptual environment. In this chapter, we shall describe some of the properties of the manifold of perceived colors and some of the determinants of color perception.

Modes of Appearance of Colors

Let us begin with the phenomenology of colors. The ways we see colors, their *modes of appearance*, can be ordered into classes upon which most investigators have agreed.

Film Colors. Look at the sky on a clear day. Or, indoors, roll up a piece of paper into a tube and through the opening of the tube view a colored object. You will experience an expanse or *film* of color, spread out in two dimensions only, without depth and lacking the characteristics of an object. It is this type of color which we usually study in determining the relations between physical properties of the stimulus and the attributes of visual experience, as discussed in Chapter 6. It is with film colors that the laws of stimulus mixture are derived. In everyday experience, film colors are rare as compared with surface colors.

Surface Colors or Object Colors. The book in front of you

is red, the desk is brown, the pencil is yellow. These colors are surface colors. They are perceived not as expanses of sheer color, but as *colors of objects*, inherent in their surface, whatever the shape of the surface. Take the red book in front of you and put it on a poorly lit corner of the desk. You perceive a red book in poor illumination. Now, take the same red book and put it directly under the light shed by the desk lamp. The perception is of a red book in bright illumination. One of the most important characteristics of surface color is that it may be perceived as distinct from the illumination falling on it. Our ability to respond differentially to light reflected from objects and to the illumination falling on these objects is invaluable in building up a stable and coherent world of colors.

Bulky Color. Look at an object through a bowl or bottle of colored or cloudy liquid. The color of the liquid in the vessel will appear to extend in three dimensions. The mode of appearance of the color is bulky. In order to have the experience of bulky color, it is necessary that you see objects in or beyond the volume of color; otherwise, the voluminousness of the color disappears.

Transparent, Lustrous, and Luminous Colors. In addition to spatial characteristics, colors have other qualitative differences. Some colors allow you to see through and beyond them, i.e., they are *transparent*. Film colors and surface colors may be transparent, bulky colors always are. With some surfaces we see not only surface colors but also a sheen of reflected light which appears to belong to the surface of the object and yet not be part of its surface color. Metallic and silken surfaces are known for their *luster*, as is also the surface of moving water with light playing upon it. Finally, filmy and bulky colors which are brighter than their surrounds may appear *luminous*. For example, if in an otherwise darkened room, we focus a beam of light on one side of a sheet of milk glass, the other side of the glass appears as a luminous film. A color is seen as luminous provided it is brighter than white in the same illumination.

The Color of Objects

Microstructure. Let us turn our attention to surface colors which are the colors of the majority of objects in our daily environment. These objects present a well-defined surface in which color is localized no matter what the gross spatial irregularities of the sur-

face may be. These surfaces are not smooth expanses of color but rather are endowed with microstructure or *grain*, i.e., slight irregularities in the surface. The presence of microstructure is conducive to the perception of colors as object or surface colors.

Brightness Differences. The appearance of surface colors is also a function of brightness differences in adjacent areas. This fact may be demonstrated by the following experimental procedure. A circular spot of light is projected on a screen. This spot appears luminous. If a ring of light of different intensity is projected around the circle, the two illuminated areas are perceived as surface colors. The brighter one appears white; the darkness of the gray of the other area depends on the ratio of the intensities of the projected lights. It is noteworthy that white is not a matter of the intensity of light alone. White, like black and gray, refers to a surface color and not to a specific degree of brightness.

Albedo and Illumination. Whenever a surface color is perceived, there are two levels of light intensity present in the situation. First of all, there is the intensity of the light falling on the object, i.e., the *illumination*. This we also refer to as the incident light. The illuminated object reflects a certain portion of the incident light. The ratio of reflected light to incident light is known as the *albedo* of the object. The albedo is a physical property of the surface of the object and is independent of the amount of illumination falling on it. When light of a certain intensity falls on a white sheet of paper, a certain percentage of that light is reflected. When the absolute intensity of illumination is raised, the paper reflects a greater absolute amount of light, but the ratio of reflected to incident light, i.e., the albedo, remains constant. At a given level of illumination, the color of an object varies with the albedo of its surface. Of course, the chromatic aspect of the surface color depends on which wave lengths are absorbed and which are reflected by the surface of the object.

Color Constancy

The Puzzle of Constancy. Look at a heap of coal lying in your yard in the bright sunlight. It appears black. Then turn your gaze to a towel hanging in a deeply shaded part of the yard. It appears white. We would be puzzled, indeed, if coal did not appear black

and towels, white. Actually, however, it is the fact that they do so appear under the circumstances which is puzzling. The coal, though its albedo is low, is reflecting a great deal of light from the sun. True, the towel has a high albedo, but, hanging in the deep shade, it is reflecting only a small absolute amount of light. The intensity of the retinal image cast by the heap of coal is much greater than that cast by the towel. If the object color were a simple function of the intensity of the retinal image, the towel should look much darker than the coal.

Such is the puzzle of color constancy. Our judgments about the color of an object agree more closely with a fixed physical property of the object—the albedo—than the retinal image alone should allow.

Illumination and Color. Somehow the organism can take into account the amount of incident light in responding to the amount of light reflected by the object. In other words, in judging the color of an object, perception discounts, at least in part, the factor of illumination. As a result, the judged difference between two colors corresponds more closely to the difference between their albedos than to the absolute difference in amounts of light reflected.

How can illumination be taken into account? Suppose there are in our visual field two objects under a given level of illumination, I_{low}. The two objects are viewed against a common homogeneous background. One of the objects is seen as black, the other as white, and the background as gray. This situation is schematically represented in Fig. 42a. The two objects and the background differ from each other only with respect to their albedos. The object seen as black has the lowest albedo, the object seen as white, the highest, and the gray background, an intermediate one. Since the incident light is the same for each of the three surfaces, and since their albedos are different, different amounts of light are reflected from them. As represented in Fig. 42, let i_1, i_2, and i_3 stand for the respective amounts of light reflected from the three surfaces. Let us express the ratio of the intensity of light reflected from the white surface to the intensity of light reflected from the gray surface as $i_1/i_2 = m$, and the corresponding ratio for the black object, $i_3/i_2 = n$.

Now let us increase the general level of illumination from I_{low} to I_{medium} (Fig. 42b). The observer still sees the three surfaces as white, gray, and black. However, since the albedos have not changed, i_1, i_2,

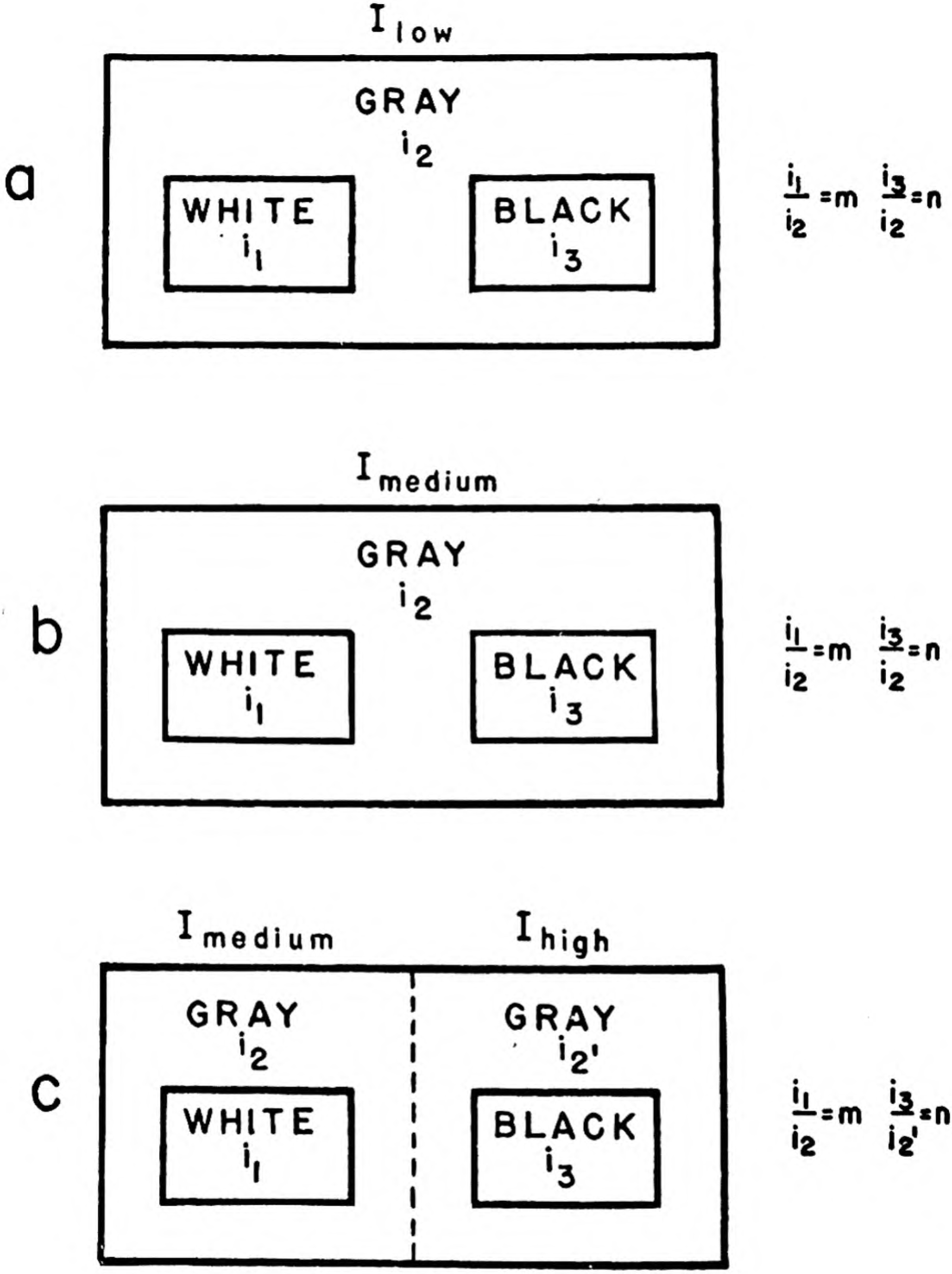

I low
GRAY
i2
a
WHITE
i1
BLACK
i3
i1/i2 = m i3/i2 = n
I medium
GRAY
i2
b
WHITE
i1
BLACK
i3
i1/i2 = m i3/i2 = n
I medium
I high
GRAY
i2
GRAY
i2'
c
WHITE
i1
BLACK
i3
i1/i2 = m i3/i2' = n

Fig. 42.

and i_3 have increased proportionately to the increase in the general level of illumination. Since the general level of illumination has been multiplied by a given factor, the amounts of reflected light have each been multiplied by the same factor. What about m and n? They must necessarily remain the same, since all three intensities have increased proportionately.

As a final change, let us keep the white object and its immediate surround at the medium level, I_{medium}, but raise the level of illumination falling on the black object and its immediate surround to I_{high} (Fig. 42c). Despite this difference in illumination, we still see on the one side a white object on a gray ground and on the other side, a black object. Why? Because m and n have not changed! Of course m has not changed since i_1 and i_2 have not changed. Now, on the right-hand side of Fig. 42c, we denote the reflected light from the gray surround which is under high illumination by i_2' to distinguish it from i_2 on the left which is under medium illumination. As compared with the situation in Fig. 42b, i_3 and i_2' constitute *proportionate* increases in reflected light, and, therefore, n, the value of their ratio, is the same as in Figs. 42a and 42b. The ratio of light reflected from an object to the light reflected from its background remains constant in spite of changes in illumination. The situation depicted in Fig. 42c is analogous to the coal in the sun and the towel in the shade, and it is clear that the constancy of these ratios greatly helps in maintaining the constancy of object colors under different illumination.

In general, viewing conditions are not as simple as those chosen for our example. Nevertheless, it remains true that intensity ratios of the amounts of light reflected from adjacent surfaces remain invariant under changed illumination. Thus, considerable color constancy obtains even under more complex conditions. Actually, diversity of the visual field, that is, the presence of several surfaces with distinct object colors and well organized spatially, is favorable to color constancy. A rather homogeneous visual field may appear black, white, or gray depending on the degree of illumination. As surfaces with different albedo values are introduced into the field, the observer is provided with reference values which establish definite relationships among the various object colors.

The dependence of color constancy on the ratios of light reflected

from adjacent surfaces is pointed up by the following experiment. In an otherwise darkened room, a bright beam of light is directed at a black disk (low albedo). This beam of light illuminates the disk and nothing else. An observer viewing the disk reports it as white. If the experimenter holds a small strip of white paper (high albedo) in front of the disk and in the beam of light, the observer sees a highly illuminated *black* disk and a white piece of paper. As soon as the strip of paper is removed, the disk again appears as white. The presence of the white strip of paper serves to establish that ratio of light intensities reflected from adjacent surfaces required for the perception of the white and the black. Without the white paper this ratio is established by the illuminated disk and the dark background of the room.

Procedures for Measuring Color Constancy. The essentials of a typical color constancy experiment are contained in the example represented by Fig. 42c. Typically, two surfaces under different illuminations are compared and equated with respect to their object color. To accomplish this, the albedo of one of the surfaces is varied until the perceived object colors of the two surfaces are the same. For example, a color wheel with a gray disk (the standard) is rotated under poor illumination. Another color wheel, with black and white disks interleaved on it, is rotated in direct, good illumination. The proportions of black and white are adjusted until the resulting gray is as near as can be to that of the standard. Most observers agree that a perfect match cannot be made. The difference in illumination is, of course, perceived and imparts a qualitative difference to the two disks.

It is only rarely that the albedos of the two matched disks will be the same. In other words, color constancy is usually not complete. Typically, the disk in high illumination is set at a lower albedo value than the disk in low illumination, and thus reflects a smaller fraction of the incident light. In spite of its lower albedo, the highly illuminated disk still sends a greater absolute amount of light to the eye than the disk in shadow. The match neither equates the albedos nor does it equate the absolute amounts of light reflected.

A match based on the absolute amounts of light reflected from the two disks *can*, however, be obtained under the following conditions: A screen with a viewing hole is placed between the observer and

the disks. The screen is so arranged that the observer can see only a patch of each of the two disks. This screen is called a *reduction screen* because it reduces the visible field to two patches with a common surround, viz., the screen. When the two test objects are provided with a common surround, the eye can be used as a photometric instrument to match the two light intensities. It is often startling to the observer when the reduction screen is removed and he finds that he has matched as equal in brilliance what now appear to be a near-white disk and a black disk.

Let us reconsider the difference between the two viewing situations—with and without reduction screen—as regards the ratios of intensities reflected from the disks and their surrounds. With the reduction screen, an equation of the two light intensities insures that the ratio of light reflected from the disk to the light reflected from the surround is the same for both test objects. When the reduction screen is removed, we do not alter the intensities of light reflected from the two disks. However, the intensities of light reflected from their surrounds are now considerably different. Therefore, the ratios in question have changed and the disk in high illumination now appears black. A perfect match with respect to object color would now require that we increase the intensity of light from the disk in high illumination (i.e., increase its albedo) so that the two ratios would again be equal. Actually, color constancy is not complete and the colors of the two objects will appear the same before the ratios have reached equality. The match is a compromise between equal light intensities and equal albedos.

A Quantitative Index of Constancy. We can express the degree of color constancy by an index which shows where the match falls on the scale between equal stimulus intensities and equal albedos. A match based on equal stimulus intensities would mean zero constancy. A match based on exactly equal albedos would mean complete constancy. Actually, most obtained matches fall somewhere between these two extremes. Let r be the albedo of the disk under low illumination. Let a represent the albedo of the disk under high illumination required for a match *without* the reduction screen. Let p represent the albedo of the disk under high illumination required for a match *with* the reduction screen. Thus, a and p represent the albedos of the object under high illumination re-

quired to make the two types of matches. Constancy can then be expressed by means of the following index:

$$c = 100 \ \frac{a - p}{r - p}$$

If two matches (with and without reduction screen) require the same albedos for the disk under high illumination, *a* equals *p*, and there is no constancy ($c = 0\%$). If, without the reduction screen, the albedo of the disk under high illumination is the same as that of the disk under low illumination, *a* equals *r*, and there is perfect constancy ($c = 100\%$).[1]

Constancy and Learning. When the phenomena of color constancy first came under consideration, an explanation was attempted by invoking the influence of past experience. Thus, the coal was said to appear black and the towel white in spite of changes in illumination because the individual knew that coal was black and towels white and had learned to discount changes in illumination. This explanation breaks down when test objects such as disks are used, for which past experience provides us with no definite expectation as to object color.

The possibility remains open that organisms do learn to some extent to discount illumination changes in making judgments of object colors. Degree of color constancy has been investigated with subjects of different ages. Under some experimental conditions—for example, with the arrangement of two color wheels discussed above—substantial increases in percent constancy as a function of age have been reported. The principal change in the amount of color constancy appears to occur about age six. Nevertheless, considerable constancy can be obtained with very young subjects and even with animals as low in the phylogenetic scale as the fish. That color constancy can to some extent be learned, may be shown by specific training procedures. In some experiments on color con-

[1] The logarithms of the values in the formula are often used and *c* is obtained from:

$$\frac{(\log a - \log p)}{(\log r - \log p)}$$

Since brilliance tends to vary in proportion to the logarithm of the physical stimulus, this formula has often been found convenient in determinations of color constancy.

stancy, bimodal distributions of color matches were initially obtained. The matches made by some subjects came close to an equality of albedos. Matches made by other subjects showed little constancy. Use of explicit training instructions made it possible to change the matching responses of both groups, in one case, decreasing, and in the other, increasing the average percent constancy. An interesting observation relevant here is the fact that some artists who try to represent faithfully the stimulus values of objects were found to show below average color constancy. Although learning to discount the effect of illumination is thus a demonstrable determinant of color constancy, it is certainly not the only one, and probably not the principal one.

Color Contrast

The importance of the relation among adjacent areas for the perception of color has become clear in our discussion of color constancy. Another group of visual phenomena, those of *color contrast*, are produced by changes in the relation of adjacent areas in the visual field. We may define contrast as changes in the brilliance and/or hue of a *test region* due to the characteristics of an *inducing region*. When we study the contrast between adjacent areas, both the test region and the inducing region are simultaneously present in the visual field. In such cases we speak of *simultaneous contrast*. Another type of contrast results from the successive stimulation of the same retinal area by different test patches and surrounds. The afterimage is one example of *successive contrast*. In this section we shall confine our discussion to simultaneous contrast.

Achromatic Contrast

If we cut two small test patches from a sheet of light gray paper and put one on a white background and the other on a black background, the two gray patches will no longer look alike. The one placed on a white background will look darker than its twin on a black ground. The two inducing regions, one white and one black, have affected the perceived color of each test patch. The change in the color of the test patch is away from the color of the inducing area. White darkens the gray and black lightens it.

The degree of achromatic contrast is a function of several determinants which can be experimentally varied.

Texture and Contours. The more apparent the texture or microstructure of the test patch and surround, the less the contrast obtained. It is for this reason that contrast is more marked when a sheet of thin transparent tissue paper is spread over the induced and inducing regions. Another way in which the texture may be lessened and contrast enhanced is by use of color wheels. If the gray test patch is a small disk at the center of the color wheel with a large ring of white surrounding it, rotation of the wheel washes out the texture and increases the contrast effect.

Related to the influence of texture is the influence of the sharpness of contours. If the induced and inducing areas are separated by sharply defined boundaries so that the two regions have the characteristics of object colors, the degree of contrast is diminished.

Brightness Difference Between Inducing and Test Areas. A large difference in brightness between the inducing and the test areas results in greater contrast than is obtained with a smaller brightness difference. A medium gray is darkened more by a white than by a light gray. In general, the greater the photometric difference in brightness between two adjacent areas, the greater the resulting contrast.

Size of Test Patch and Surround. The greatest contrast effect is obtained if the test area is small and the inducing area is large. Increases in the size of the test patch and decreases in the size of the inducing area both serve to reduce contrast.

Gradient of Contrast. The contrast effect is most marked where the test patch makes contact with the inducing area. The contrast diminishes with the distance from the edge of the test patch.

Black. The perception of black is a phenomenon related to achromatic contrast. Black is a surface color and not the absence of stimulation. Black surfaces typically have a low albedo value. The surround of such a surface of low albedo is an important determinant of the degree of blackness that is seen. In high illumination, a low albedo surface surrounded by another surface with low albedo will not appear very black. Under what conditions do we see a good black? The blackest black can be obtained by employing both successive and simultaneous contrast. To obtain suc-

cessive contrast, we steadily fixate a brightly illuminated white surface with a black surround, and then turn our gaze to a velvety black surface (same size and shape as the white one) surrounded by a white field under low illumination. The black surface with a white surround under low illumination creates simultaneous contrast. It is by utilizing the two types of contrast that the blackest black is seen.

Chromatic Contrast

Chromatic colors, as well as achromatic ones, can give rise to simultaneous contrast. As a first demonstration, let us take a gray test patch and put it on a well saturated blue field. The gray patch takes on a distinctly yellowish tinge which is most marked at the edge of the test patch adjacent to the blue ground. The blue field has induced its complementary, yellow, onto the neutral test patch. Although chromatic contrast can best be demonstrated with a neutral test patch, simultaneous contrast effects can also be obtained if both the inducing field and the test field are chromatic. For example, when we have a yellow surrounded by a red, the yellow will take on a bluish green (complementary of red) and the red will take on a bluish tinge (complementary of yellow).

Chromatic contrast, like achromatic contrast, is a joint function of several determinants.

Saturation of Test Patch and Surround. Chromatic contrast is maximal if the inducing field is richly saturated and the test patch is poorly saturated. A test field of zero saturation, i.e., a gray, results in the highest degree of chromatic contrast.

Brightness Difference Between Inducing and Test Areas. Chromatic contrast is greatest when the inducing and test areas are of equal brightness. A brightness difference between the two areas serves to reduce the contrast as compared with the condition of equal brightness.

Texture and Contours. Chromatic contrast, like achromatic contrast, is resisted by areas that have clear surface color or marked texture. Again, the presence of well-marked contours decreases the amount of contrast as the following demonstration shows.

In Fig. 43, we see a field divided in half with one side red and the other green. A gray ring lies in the center of the figure so that

half is on red and half is on green. When no dividing line is present, i.e., when there is a complete ring, the ring is seen as gray and is not tinged due to chromatic contrast. When the dividing line is drawn, as shown in the figure, the semicircular ring on the green field takes on a reddish tinge and the other half of the ring, which lies on the red field, takes on a bluish-green tinge. According to one explanation of this phenomenon, the complete ring is a well-

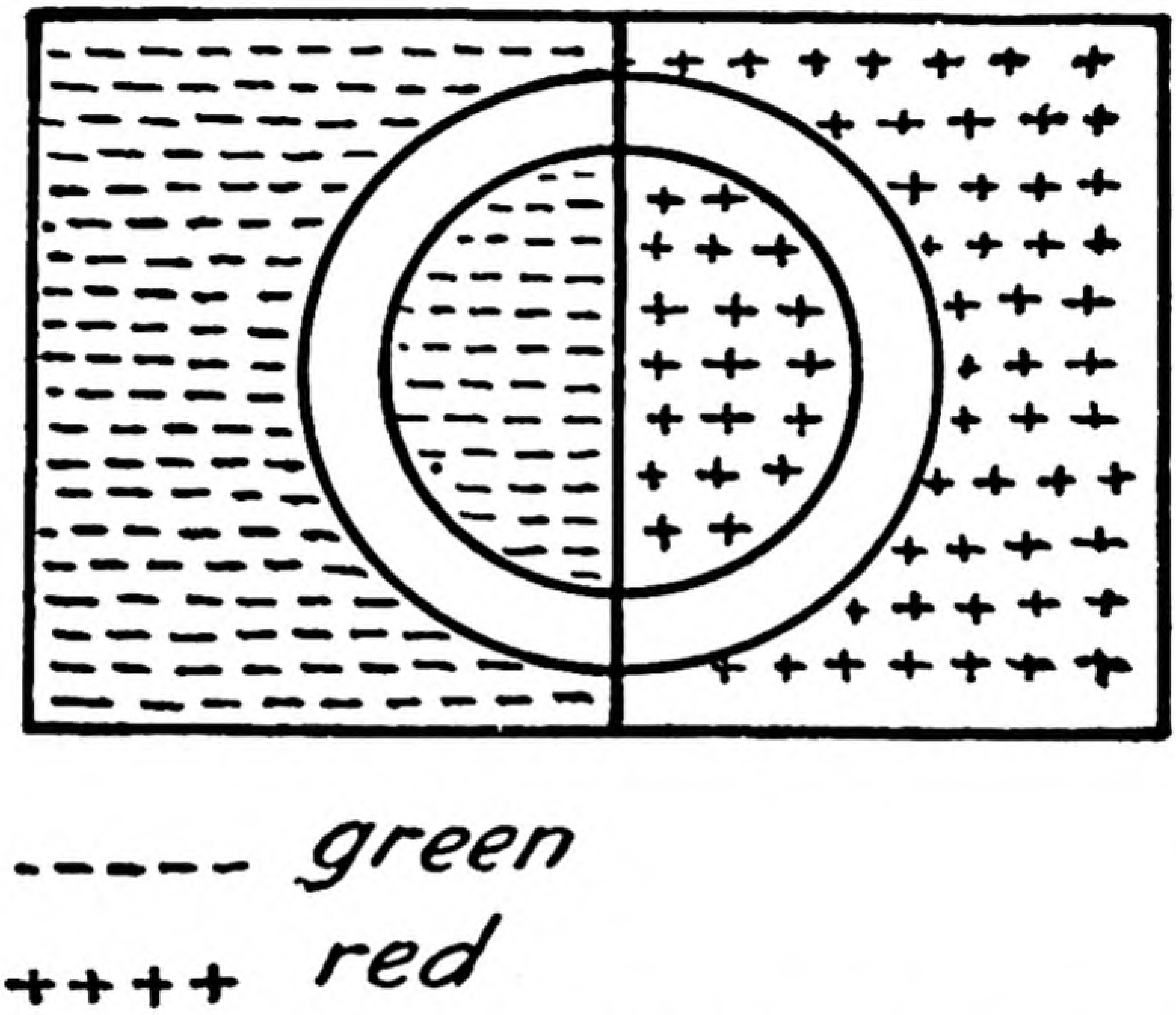


Fig. 43. The Dependence of Chromatic Contrast on Contour. For explanation, see text. (From K. Koffka, *Principles of Gestalt psychology*, 1935, p. 134, by permission of Harcourt, Brace and Company, Inc.)

organized unit which resists change. The dividing line disrupts the spatial unity of the ring, and the contrast effect appears.

Size of Test Patch and Surround. As far as size relations between the two areas are concerned, chromatic contrast is greatest if the test patch is small and if the inducing field is large.

Gradient of Contrast. One final parallel between chromatic and achromatic contrast may be pointed out. For a given viewing condition, chromatic contrast is greatest at the margin of the test area where it makes contact with the inducing field.

Experiment VIII
Color Constancy

Purpose. To demonstrate and measure color constancy with achromatic surfaces.

Materials. Two color wheels, white and black color disks, a shadow caster, a reduction screen, and a lamp are the required materials. The shadow caster is merely a rigid, flat, upright surface which is placed so that it casts a shadow on a color wheel.

Procedure. On a table set against a wall, two color wheels are mounted side by side. One of the wheels is illuminated by the lamp. The shadow caster is placed between the two wheels so that it casts a shadow on the second wheel. On the wheel in the shadow, black and white disks are interleaved: 40 percent white and 60 percent black. This ratio remains fixed for a given set of observations. Black and white disks are also interleaved on the illuminated wheel, but the proportions on this wheel are to be adjusted to obtain a match.

In the first part of the experiment, the subject's task is to match the black-white mixtures on the two wheels with respect to grayness (object color). The proportions on the illuminated disk are changed as indicated by the subject until a satisfactory match is obtained. During this match, the subject has full view of the two wheels and the adjacent parts of the table and wall. The amounts of white and black required for the match are recorded. Several matches should be obtained under these conditions. For half of these determinations, the illuminated disk should be initially much too light and the proportion of black increased until the subject is satisfied with the match. For the other half of the determinations, start with the illuminated disk much too dark and increase the proportion of white until the match is made.

In the second part of the experiment, a reduction screen is placed in front of and near the subject's eyes. The holes in the screen should be so placed that only patches of the disks can be seen. Again, the proportions of black and white on the illuminated disk are adjusted until a match with respect to brilliance is made. Several determinations should also be made of the reduction screen match. At the end of this procedure, the reduction screen is removed and the subject again has an open view of the two color wheels.

Treatment of Results. On the basis of the matches made in the two parts of the experiment, the index of constancy is computed:

$$c = 100 \frac{a - p}{r - p}$$

where r stands for the percentage of white on the disk in low illumination; a is the percentage of white on the illuminated disk required for a match without the reduction screen; and p is the percentage of white on the illuminated disk required for the reduction screen match.

The reader may note that we have substituted percentage of white for albedo (see discussion of this index on p. 153). We are making the assumption that the black reflects no light. This assumption is not correct, since even a good black reflects some light. One degree of white reflects approximately as much light as 60 degrees of black. To simplify the computation of the constancy index we will disregard this fact.

Experiment IX

Demonstration of Achromatic and Chromatic Contrast

Purpose. To observe some phenomena of contrast.

Materials. A series of squares of gray paper graded in size from ½ inch square to 2 inches square are prepared. In addition, there are series of black, white, red, yellow, green, and blue squares. Each series of these latter squares should range in size from 3 inches square to 6 inches square. A sheet of white transparent tissue paper is also needed.

Procedure. For the demonstration of achromatic contrast, we place a white square and black square of the same size side by side. In the center of each of these inducing fields, we place a gray square, which is of the same size in both fields. Then we cover the whole area with the tissue paper. A comparison is then made of the color of the two test patches. This procedure is followed for the various combinations of sizes of test patch and inducing field. The degree of contrast observed using different size relations is noted.

For chromatic contrast, the gray test patches are placed upon chromatic inducing fields and covered with tissue paper. Various combinations of sizes of test patch and surround are again viewed. For each chromatic field and for each size combination, the observations are recorded. Compare the effect obtained with and without the use of tissue paper.

General References

Boring, E. G., Langfeld, S. H., and Weld, H. P., *Foundations of Psychology*, New York: John Wiley and Sons, 1948, Chap. 12.

Katz, D., *The world of colors*, London: Kegan, Paul, Trench, Trubner and Co., 1935.

Woodworth, R. S., *Experimental psychology*, New York: Henry Holt and Co., 1938, Chap. 24.

Modes of Appearance of Colors

Katz, D., *The world of colors*, London: Kegan, Paul, Trench, Trubner and Co., 1935, Part I.

Woodworth, R. S., *Experimental psychology*, New York: Henry Holt and Co., 1938, Chap. 22.

Color of Objects and Color Constancy

Koffka, K., *Principles of Gestalt psychology*, New York: Harcourt, Brace and Co., 1935, Chap. 6.

MacLeod, R. B., An experimental investigation of brightness constancy, *Arch. Psychol.*, 1932, No. 135.

Sheehan, M. A., A study of individual consistency in phenomenal constancy, *Arch. Psychol.*, 1938, No. 222.

Thouless, R. H., Phenomenal regression to the real object, *Brit. J. Psychol.*, 1931, *21*:339-359; 1931, *22*:1-30.

Wallach, H., Brightness constancy and the nature of achromatic colors, *J. Exper. Psychol.*, 1948, *38*:310-324.

Color Contrast

Boring, E. G., *Sensation and perception in the history of experimental psychology*, New York: D. Appleton-Century Co., 1942, pp. 165-171.

Koffka, K., *Principles of Gestalt psychology*, New York: Harcourt, Brace and Co., 1935, pp. 133-138.

Woodworth, R. S., *Experimental psychology*, New York: Henry Holt and Co., 1938, Chap. 22.

»»» 8 «««

PERCEPTION OF FORM

THE individual continually adjusts himself to the objects in his environment. He responds, often with amazing accuracy and speed, to the specific things surrounding him. All of us take the presence of these things and objects for granted. In our everyday life, we continually react to their properties—their size, their pattern, their location, and their meaning—and yet we only rarely make explicit judgments about these properties. As one psychologist has expressed it, speaking about things seen in the environment, these things "simply *were there outside*, and . . . I had no suspicion whatever of their being the effects of something else upon 'me.'"[1]

The perception of things and objects poses a complicated problem for the psychologist. Obviously, the visual perception of an object is impossible without an optical image on the retina. On the retina, the object is represented by a continually changing pattern of discrete stimulations. This pattern may best be thought of as composed of gradients of stimulation. Ordinarily, every part of the retina is stimulated to a greater or lesser extent; different objects in the visual field cause different levels of excitation or gradients. How is it that these highly complex gradients of stimulation on the retina give rise to those processes which finally result in the perception of a relatively stable and richly varied environment of things and objects? That, in essence, is the problem to which the study of the perception of form addresses itself.

Figure and Ground

When our environment produces perfect homogeneous stimulation of the visual receptor surface, we do not see objects and things. It is gradients of stimulation on the retina which produce objects

[1] W. Köhler, *Gestalt Psychology*, New York: Liveright, 1929, p. 21.

standing out from their environment. Our visual field, moreover, is not completely filled with objects. Objects and things appear as *figures* against a *ground.*

The Experience of Figure and Ground

The fundamental distinction between figure and ground is well borne out by experience. Our visual field readily divides itself into

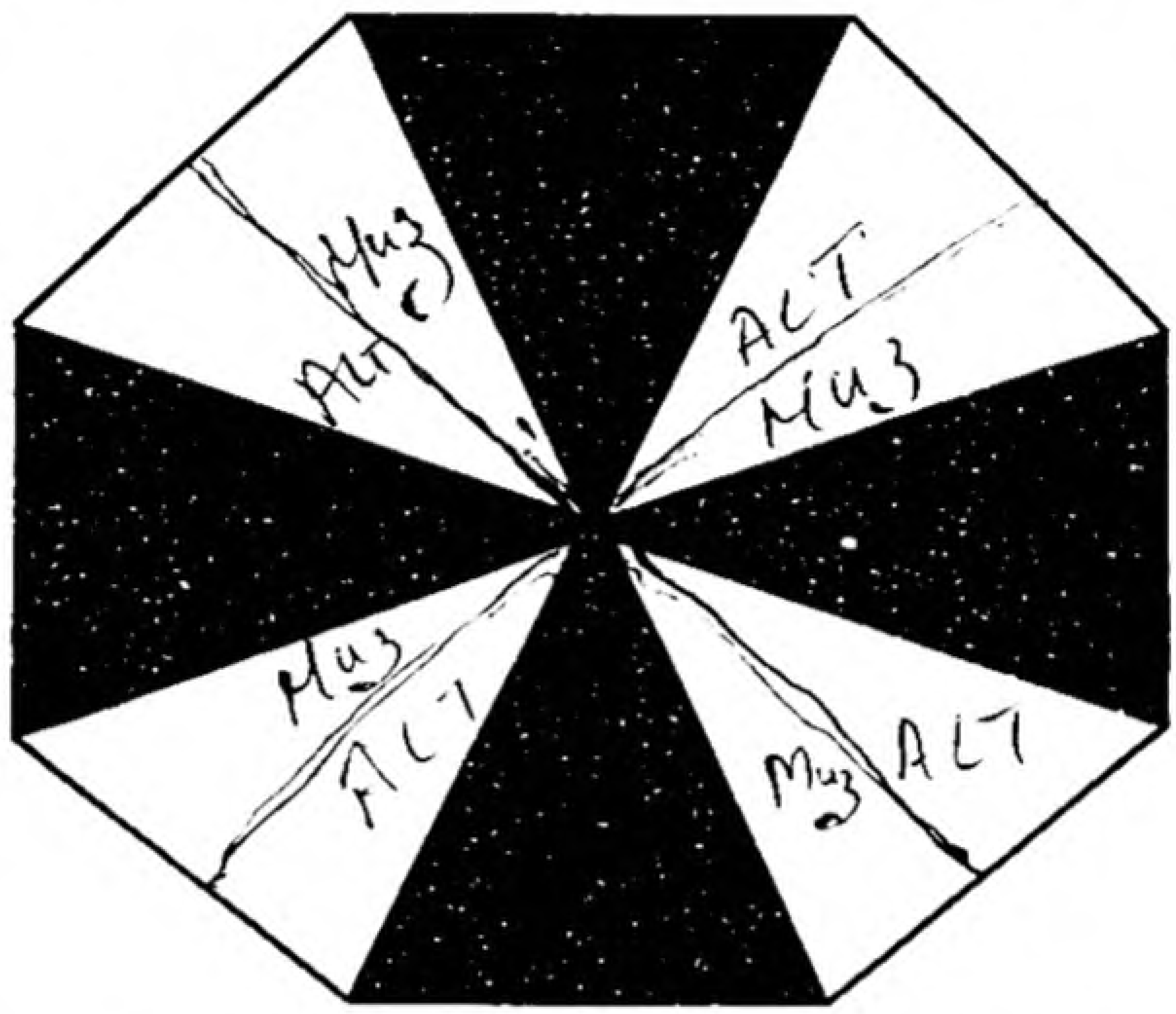


FIG. 44. Reversal of Figure and Ground. The figure may be seen as a black Maltese cross against a white background or as a white propeller against a black background. (From George W. Hartmann, *Gestalt psychology,* p. 25. Copyright 1935, The Ronald Press Company.)

a broadly expansive ground and a sharply delineated figure. A car may stand out as an impressive figure against a vague landscape which constitutes the ground, and so may an airplane seen against the sky. Such properties of figures and grounds may best be illustrated with the aid of simple figures.

Let us look at Fig. 44. This figure consists of black and white surfaces. The black surfaces may be seen as a Maltese cross against a white background, or the white surfaces may form a white propeller against a black background. For most observers, both percep-

tions are possible and alternate in time. Whichever of these two figures stands out has certain properties which the reader may easily verify.

1. The figure stands out in front, and the ground extends behind it.
2. The figure has more form or structure than the ground. The black Maltese cross, when seen as the figure, is clearly outlined against a white background.
3. Most observers agree that the figure is more impressive, lively, and substantial than the ground.

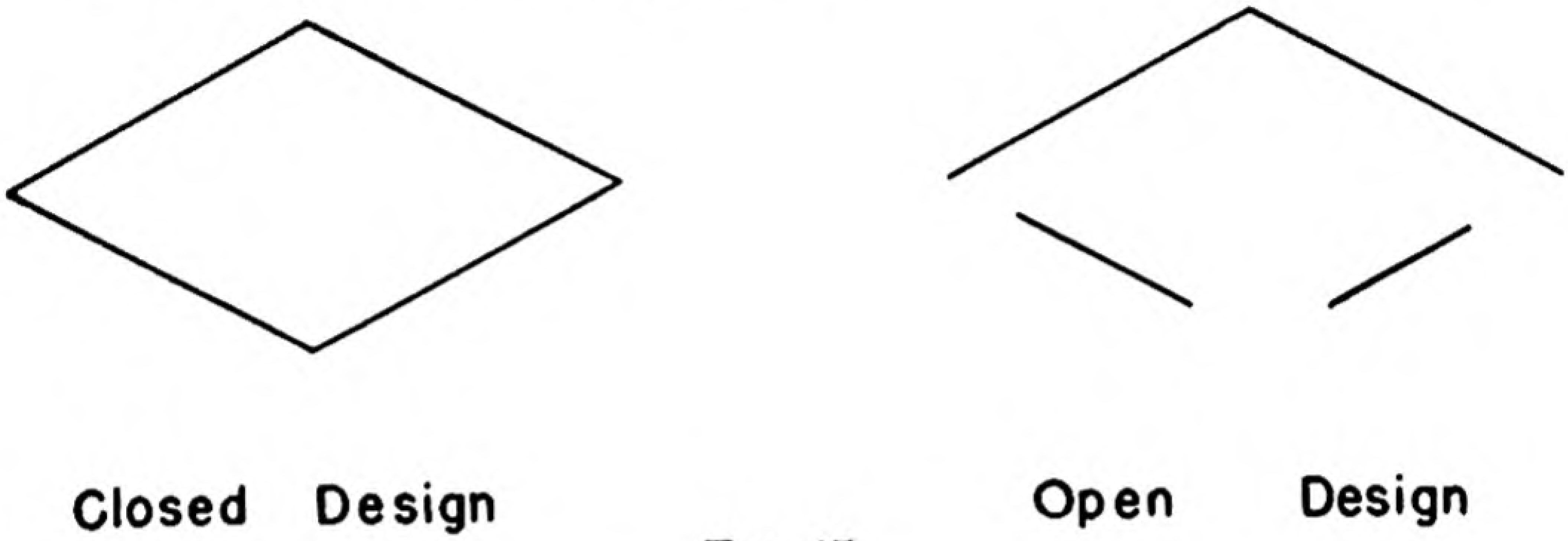


FIG. 45.

4. Thus, the figure appears more readily as a thing or object than does the ground.

Some Determinants of Figure-Ground Segregation

A full description of the processes which lead to the segregation of figure from ground is not yet possible. Some determinants, however, have been experimentally established.

Size. When there are two distinct areas lying one within the other, the smaller of the two tends to be seen as the figure. The ready reversal of figure and ground characteristic of Fig. 44 is due in part to the fact that the black and white surfaces are equal in area.

Completeness of the Figure. A closed design is more readily seen as a figure than an open one. Once seen as a figure, the closed design shows the various properties of a figure to a more striking degree. Fig. 45 illustrates this point.

Brightness Difference. Brightness difference between figure and ground facilitates the segregation. We are dealing here with one of the most important determinants of figure-ground segregation.

If the surfaces in Fig. 44 were of only slightly different shades of gray instead of being black and white, the figure would not stand out as clearly from the ground as it does.

Hue Difference. Difference in hue facilitates figure-ground formation. Certain hues are more effective than others in producing a figure. A red field, for example, will stand out more clearly from a gray background than will a blue one.

Length of Exposure Time. The clearness with which a figure is perceived against a ground depends in part on the length of time for which an observer may view the stimulus configuration. After the exposure of a stimulus for 10 milliseconds, observers usually report the first indication of a figure-ground relationship. As the time of exposure is lengthened, other properties emerge. Continuous and enclosed contours are recognized; the figure begins to protrude and the ground to recede. In general, the longer the exposure time, the better the figure-ground relationship is articulated.

These factors, singly or in combination, do not fully determine the segregation of figure from ground. It is true that they describe important and reliable conditions for the formation of visual figures. Such factors as attitude, previous experience, and practice, however, may either weaken or strengthen the operation of the above determinants. It must be emphasized again that the formation of a figure is the resultant of a highly complex process which cannot be exhaustively described by a few general principles.

The Functional Properties of Figure and Ground

Visual inspection shows striking differences between figure and ground, but experimental analysis is required to reveal functional differences that are not immediately apparent by mere inspection. The differences between figure and ground manifest themselves by measurable effects on other psychological variables.

Threshold for the Detection of Color. The minimum intensity required for the detection of a color varies with the nature of the field on which the color is projected. Suppose a patch of red light is thrown on a design which may be seen as either figure or ground in a complex configuration. A greater intensity of light is required for the detection of the red patch when the design is seen as a figure than when it is seen as a ground. This finding has been

interpreted to mean that a figure is more resistant to change in its perceptual character than is a ground.

Color Constancy. As we have seen, objects tend to retain their color in spite of changes in illumination. As illumination changes, the object color of a design is affected less if this design is seen as a figure than if it is seen as a ground. Again, the figure is more resistant to perceptual change.

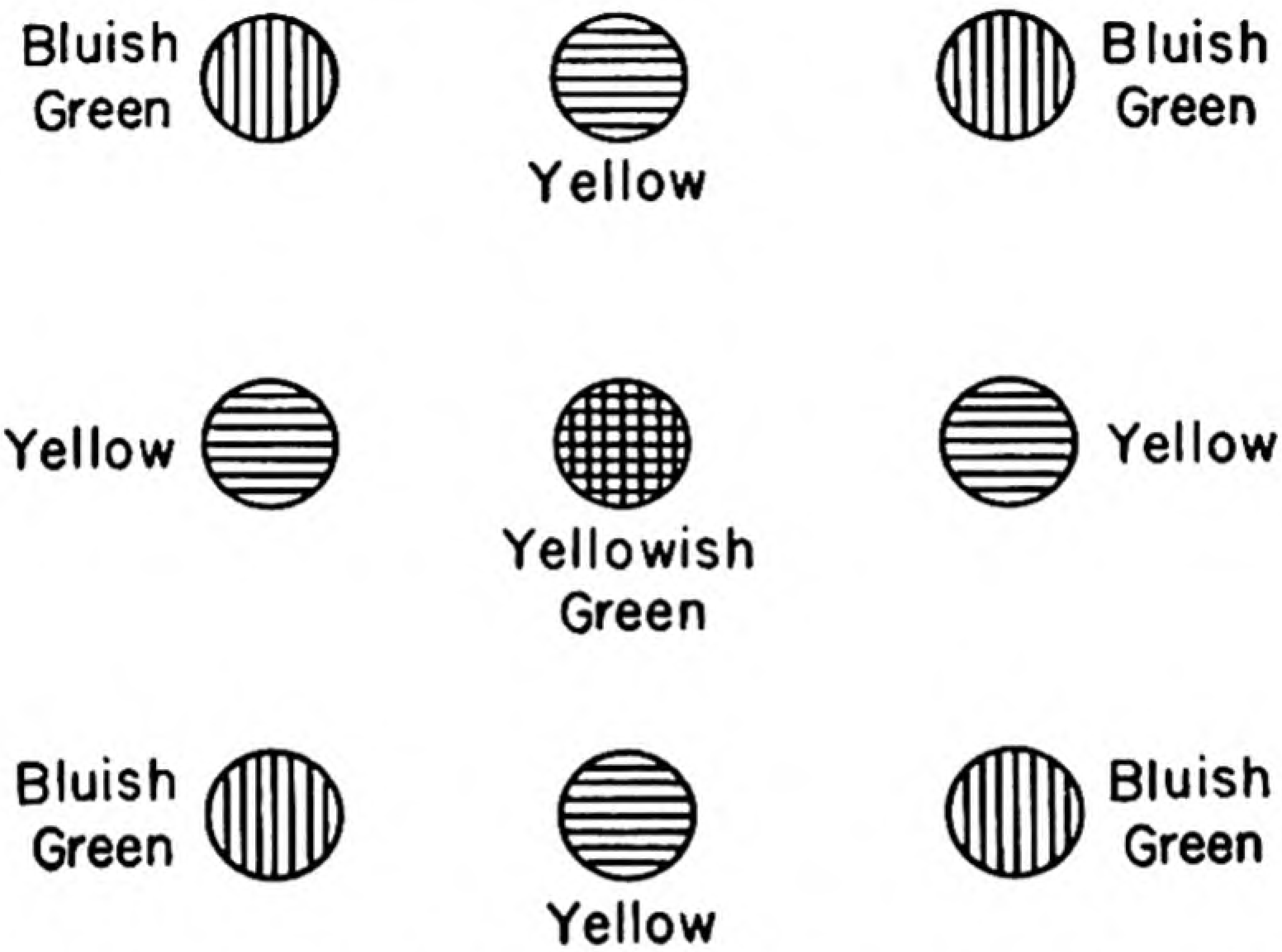


FIG. 46. The Dependence of Apparent Color on the Spatial Characteristics of the Design. The color of the center dot depends on whether the yellow plus sign or the blue-green multiplication sign is figural. (From W. D. Ellis, *A source book of Gestalt psychology*, 1938, p. 100, by permission of Routledge and Kegan Paul, Ltd. After Fuchs, 1923.)

Apparent Color. Apparent color may be influenced by the spatial characteristics of a design. Fig. 46, for example, consists of three rows of colored dots. Observers may see a yellow plus sign against a ground of blue-green dots, or a blue-green multiplication sign against a yellow ground. The apparent color of the greenish-yellow dot on the center may depend upon which of the two designs is figural: when the plus sign stands out, the center dot is predominantly yellow; when the blue-green multiplication sign stands out, the center dot is predominantly blue-green.

Persistence in Memory. Suppose that alternative figure-ground relationships are possible, as is the case in Fig. 44. If one of these alternatives is perceived at a given time, it is likely to be perceived again when the stimulus is presented at a later time. That which has been seen as figure on the first presentation will tend to be seen as figure again.

Such findings make it clear that the concepts of figure and ground refer not only to reported visual experience but also to a set of functional relationships which need to be demonstrated by experimental manipulation.

Perceptual Grouping

The segregation of figure from ground is a primary condition of perceptual organization. But our visual field does not usually consist of a single object against a homogeneous background. Ordinarily, there are many objects simultaneously present in the visual field. These objects tend to be perceived in *groups*. Such perceptual grouping is not haphazard but is governed, at least in part, by certain general principles. These principles can again best be illustrated with the aid of simple designs. Here are some of these principles:

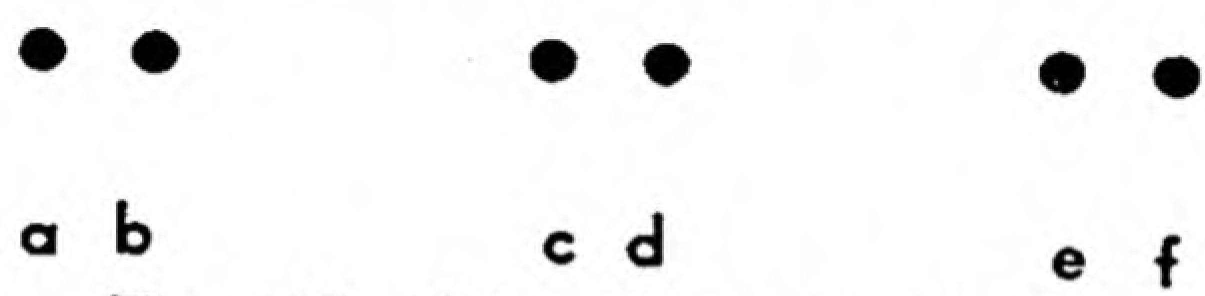


Fig. 47. The Principle of Proximity in Perceptual Grouping.

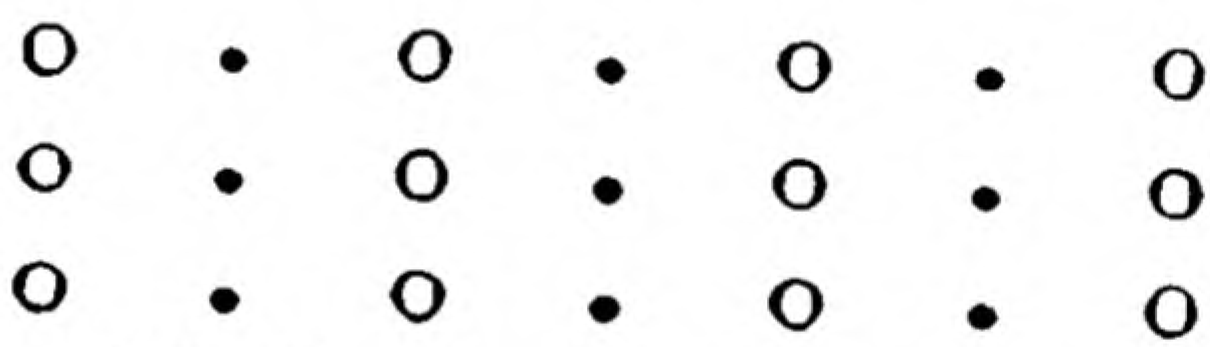


Fig. 48. The Principle of Similarity in Perceptual Grouping.

Proximity. Objects are often grouped according to the distance which separates them from each other. Objects close together tend to be seen as a group. When several groupings are possible, that

one will tend to be favored which will make the average distance among the members of the group as small as possible. This principle is illustrated in Fig. 47. We readily group dots *a* and *b*, *c* and *d*, *e* and *f*, but only with great difficulty can we achieve the groupings *b* and *c*, *d* and *e*.

Similarity. If a configuration comprises several distinct kinds of stimuli, those which are alike will tend to be grouped together. In Fig. 48, the principle of proximity would predict no grouping since all the items are equidistant. Yet, grouping clearly occurs: the like dots form one constellation and the like circles another.

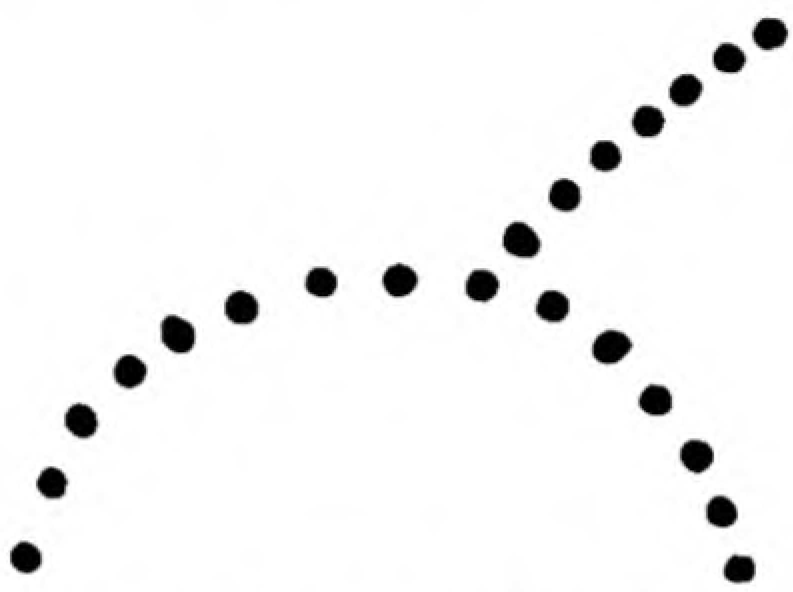


FIG. 49. Common Direction as a Principle of Perceptual Grouping.

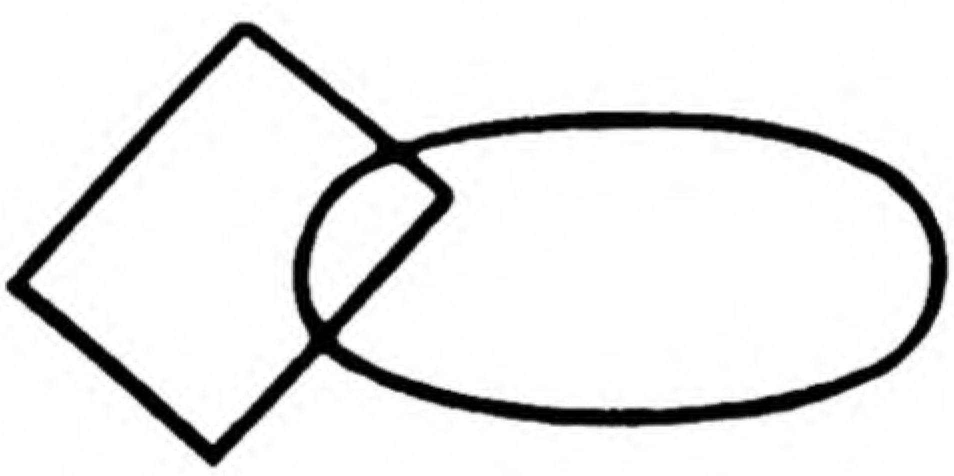


FIG. 50. An Illustration of the Principle of Closure.

Direction. When individual stimulus items fall along a straight line or along a simple curve, they will tend to be grouped together even though such grouping may be contrary to the principle of proximity. The common direction followed by a series of items is the basis of grouping. In Fig. 49, we readily segregate a straight line and a semicircle as forming two distinct groupings. Clearly, however, several dots belonging to the straight line are nearer to some dots on the semicircle than they are to some of the dots on

the straight line. In this case, then, the principle of direction overrides the principle of proximity.

Closure. In many cases, grouping tends to favor simple closed designs. This principle of closure is illustrated in Fig. 50. Here we see two overlapping completed figures instead of seeing three enclosed areas. If the area of overlap were seen as distinct, then both the "ellipse" and the "rectangle" would be broken and incomplete. Instead, the overlapping area is perceived as belonging to both, giving rise to two simple and completed figures.

The principles listed above are by no means exhaustive. Again, such factors as attitude and past experience are often operative. Moreover, as we have seen, these determinants may operate in conjunction or in opposition to each other. Proximity may at times outweigh similarity, and direction may override proximity. Quantification of these factors has not as yet been achieved.

Form Constancy

We have already seen (in connection with color constancy) how perceptual experience tends to be more stable than could be predicted from a simple analysis of the physical stimulation alone. Changes in illumination do not result in corresponding proportional changes in perceived color. The perception of form exhibits similar stability. The geometrical form of the retinal image constantly changes as we move about with respect to the object. The image projected on the retina by a circular plate lying before us on the table is actually elliptical in form. The exact form of this ellipse, moreover, varies with every movement of the head. Yet the plate does not vary in apparent shape—it remains circular. This phenomenon illustrates another kind of constancy—form constancy.

Geometrical Illusions

Perception *seems* to mirror the properties of the object with amazing fidelity. To us, our perceptual experience appears to represent the external world with a high degree of accuracy. For a long time the simple view prevailed that the retinal image is a faithful reflection of the physical stimulus, and that the perception, in turn, is an accurate copy of this retinal image. In this way, the accuracy of perception was thought to be explained. To those who held this

view, the optical illusions presented a problem of special interest and challenge. For here were experiences which obviously failed to represent accurately the physical object and its image on the retina. The optical illusions were treated as puzzling exceptions to the general laws of perception, and special principles were sought to explain them.

Modern thought about perception necessarily rejects this naïve view of the illusions. One thing stands out clearly from what we know about perceptual processes: visual perception is not a copy of the retinal image. The retinal image is only a first step in the perceptual processes; it is, as it were, the raw material out of which visual experience is shaped. Thinking back to such a phenomenon as color constancy, we also recall that the perception of an object depends not only on its retinal image but on the stimulation of

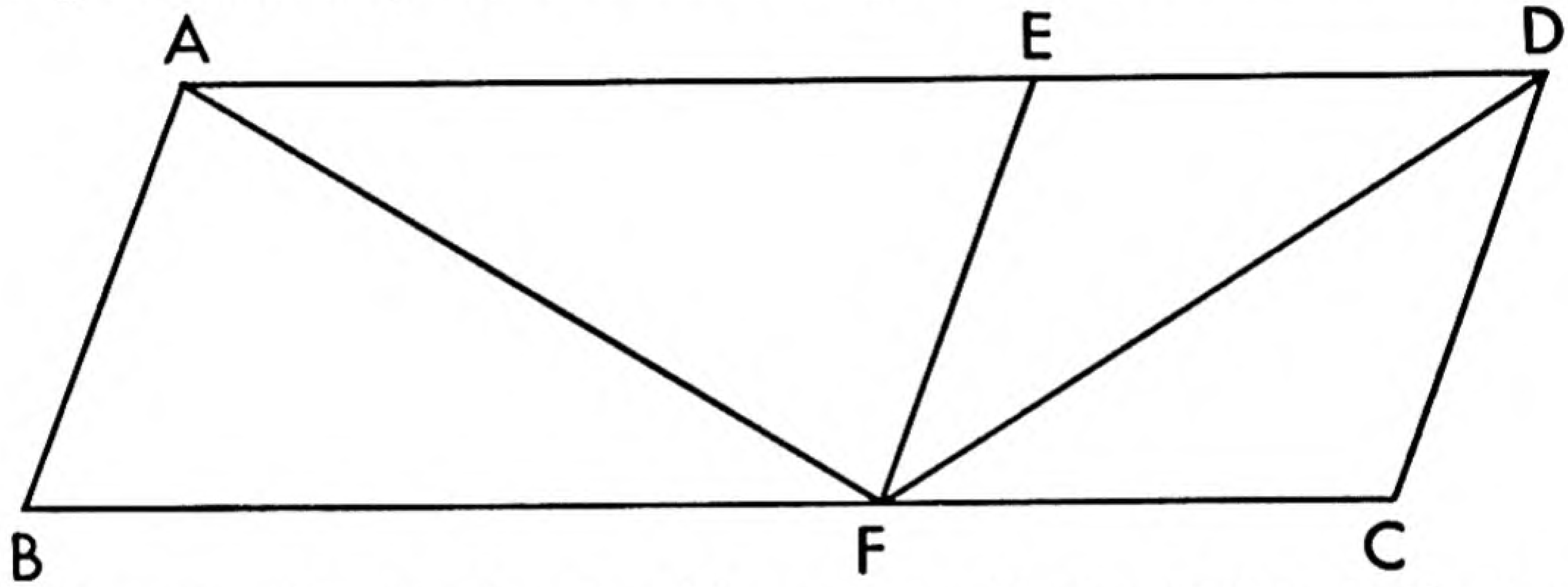


FIG. 51. The Sanders Parallelogram—An Illusion of Extent. The two diagonals are equal in length. (From George W. Hartmann, *Gestalt psychology,* 1938, p. 85. Copyright 1935, The Ronald Press Company.)

the total visual receptor surface. An illusion, then, presents no more and no less of a problem than a so-called accurate perception. In both cases, there is an image on the retina which gives rise to a further series of processes resulting in the final reportable perception. Boring wrote, "In the sense that perception is normally dependent upon subjective factors as well as upon the stimulus, all perception is 'illusory' in so far as it does not precisely mirror the stimulus. In this broad sense the term *illusion* becomes practically meaningless."[2]

[2] E. G. Boring, *Sensation and perception in the history of experimental psychology.* New York: D. Appleton-Century, 1942, pp. 238 f.

Examples of Illusions

The study of illusions, however, still has a certain value. Illusions serve as rather striking illustrations of the way in which "subjective" factors, as well as the physical properties of the stimulus, determine perceptual experience.

Two general kinds of illusions have often been distinguished: illusions of *extent* and illusions of *direction.*

Illusions of Extent. An illusion of extent arises when two physically equal stimuli are judged as different in length. Usually, such stimuli are parts of a complex design. Fig. 51 shows an example of such an illusion. Hard as it is to believe at first glance,

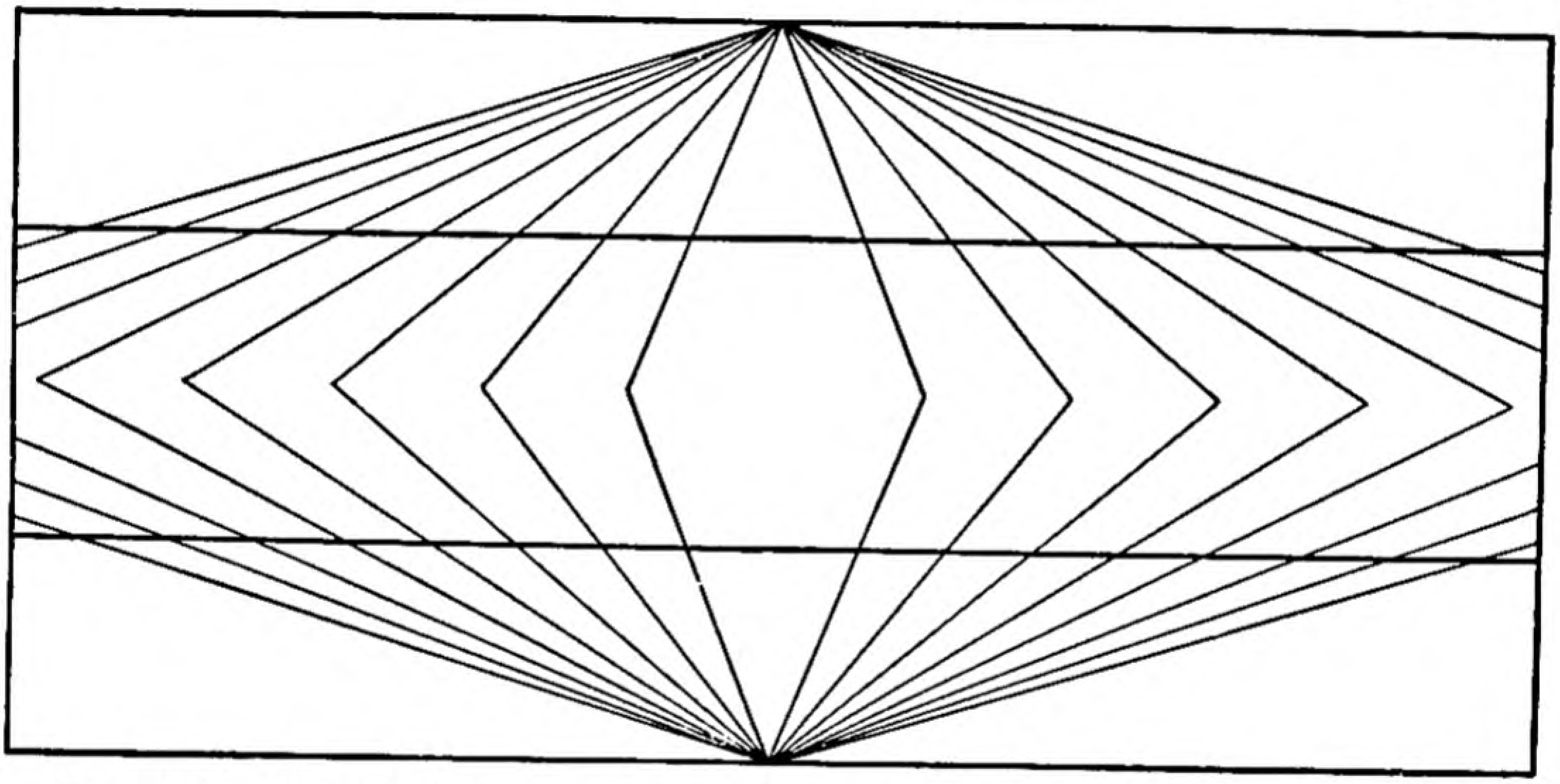


FIG. 52. The Wundt Lines—An Illusion of Direction. The two physical lines are straight and parallel. (From E. G. Boring, *Sensation and perception in the history of experimental psychology,* 1942, p. 242, by permission of Appleton-Century-Crofts, Inc.)

the two diagonals *AF* and *FD* are equal in physical length. Taken out of the parallelogram, they would, of course, appear very nearly equal.

Illusions of Direction. An illusion of direction arises when the apparent direction of a geometric form deviates from that expected on the basis of its physical direction. In Fig. 52, for example, the two horizontal physical lines are straight and parallel

to each other. Yet both of them appear to diverge toward the sides of the design. Again, if taken out of this context, they would appear straight and parallel.

Experiment X
Perceptual Grouping

The following experiment is designed to demonstrate the principles of perceptual grouping discussed above. It will serve not only to illustrate these principles but will also show how the various determinants of grouping can be made to cooperate and conflict with one another.

Purpose. To demonstrate the effects of proximity, similarity, and direction on perceptual grouping, and to measure the stability of the resulting perceptual groups.

Materials. Two kinds of colored thumbtacks, for example, blue and red ones, several sheets of cross-section paper, a wooden board, and a metronome are all the equipment required for this experiment.

Procedure. The subject is seated at a table in a well-lighted room. A wooden board is placed vertically on a table at a distance of about 2 feet from the subject. Care should be taken that movements of the experimenter or the subject do not cast shadows on the board.

A sheet of cross-section paper is tacked on the board. A design of colored thumbtacks is tacked on the cross-section paper. The subject is instructed to look for a particular figure in the design. He is also instructed to hold up his hand (or use some other convenient signal) as long as he can see that figure. A metronome is set to beat once per second so that the experimenter may tally the number of beats during which the figure is seen and during which it is not seen. The experimenter will do well to have prearranged tally sheets on which he can record (1) the nature of the design, (2) the total duration of exposure of the stimulus, and (3) the number of beats during which the critical figure was seen.

We shall now suggest some designs to illustrate the principles of proximity, similarity, and direction. Other designs should be worked out by the experimenter and subject in coöperation.

1. *Proximity.* Tacks of one color only should be used. One design consists of three columns of tacks, say, 2 inches apart. The dots in the column should be equidistant from each other, say, about ¼ inch. At the top and ¾ inch to the right and to the left of the middle column, one tack each is placed. The subject is instructed to look for the figure forming a letter T in the center. This configuration is exposed for 1 minute. After that, the design should be progressively complicated by the addition of further dots so that the principle of proximity

favors the perception of a T less and less. Each of the additional designs should be exposed for 1 minute.

2. *Similarity.* Tacks of two different colors are now to be used. A basic design consists of five rows of five tacks each. In the first configuration, the T is made readily apparent by having the top row and middle column made of one color (e.g., red) and the other tacks of the other color (e.g., blue). Again, the design is exposed for 1 minute. As before, the task of holding the T is made more and more difficult by replacing more and more of the red dots with blue dots.
3. *Proximity vs. similarity.* In this design, the principles of similarity and proximity are pitted against each other. A basic design is shown in

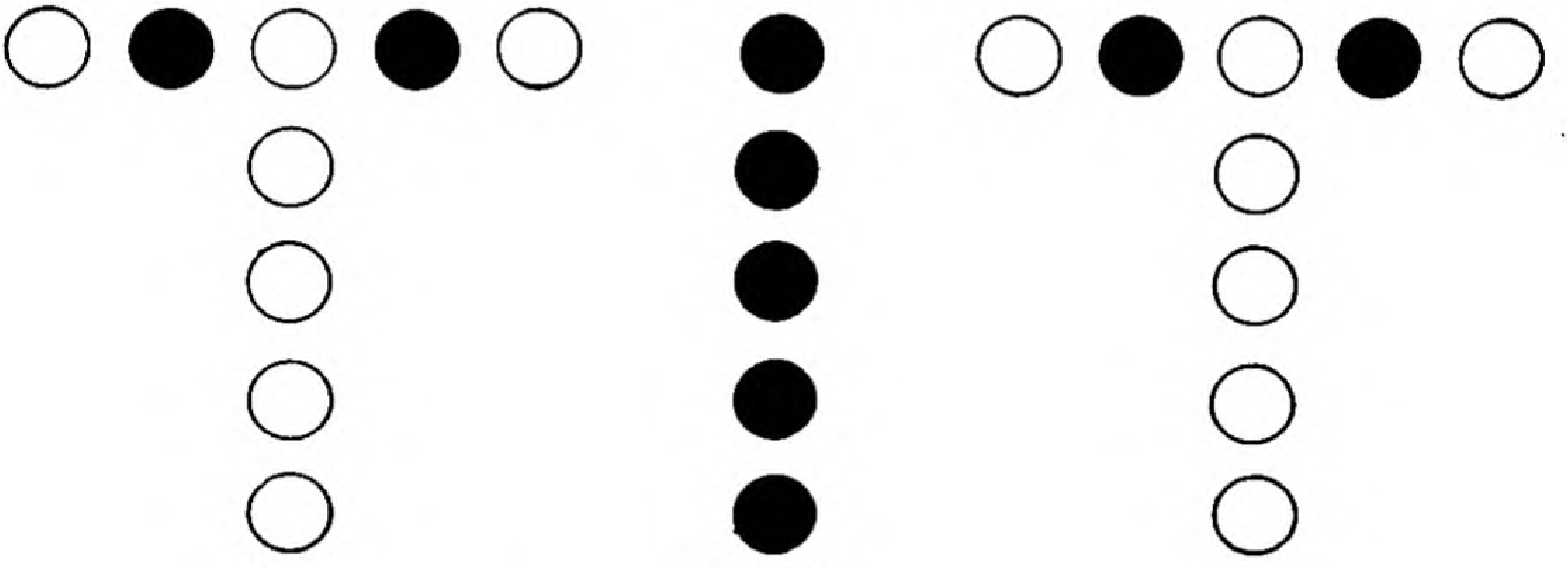


FIG. 53.

Fig. 53. Each of the two outer columns consists of five blue dots spaced ½ inch apart. The middle column is made up of five red dots similarly spaced. The columns are 2 inches apart. At the top of each of the outer columns, the arm of a T is provided by four symmetrically distributed dots, again spaced ½ inch apart. The two outer dots are blue, the two inner dots are red. The subject is instructed to look for a red T, which he would see in accordance with the principle of similarity. The principle of proximity would predict that the subject would see the two T's at the sides of the design. This design is exposed for 1 minute. Similarity can then be made even less operative by gradually replacing the outer blue tacks with red ones.
4. *Direction.* To demonstrate the effect of direction, a design of the kind reproduced in Fig. 54 or any suitable part thereof can be used. In this way, the relative stability of different lines and curves can be tested by comparing the length of time for which they can be seen by the subject. In the manner indicated above, the factor of direction can be pitted against proximity and similarity.

Treatment of Results. For each of the designs, the proportion of time during which the subject saw the figure should be computed. This is done, of course, by dividing the number of seconds during which the figure was seen by the total exposure time. Thus, an *index of stability* is obtained for each design. The stability of any one design can be compared with that of any other. These indices of stability would, of course, be somewhat different if the experiment were repeated. Small differences among these indices may arise as a result of uncontrolled factors. Before

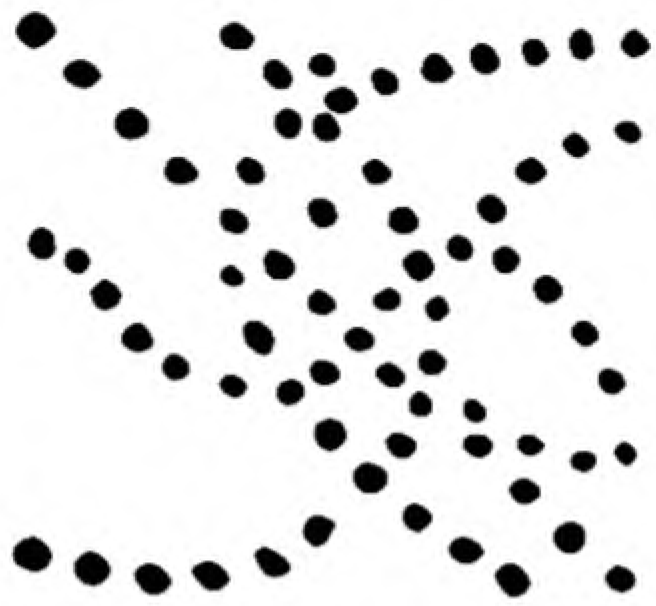


FIG. 54. A Design to Test the Effect of Direction on Perceptual Organization. (From R. S. Woodworth, *Experimental psychology*, 1938, p. 626, by permission of Henry Holt and Company, Inc.)

confidently asserting that one design is more stable than another, it would be necessary to determine how frequently a difference of a given size would occur as a result of uncontrolled factors.

EXPERIMENT XI
FORM CONSTANCY

In this experiment, form constancy is demonstrated and measured. It will be remembered that the image projected on the retina by a circular disk is often actually elliptical in form. Yet, over a considerable range of rotation, the turned circle tends to retain its apparent circularity. The following experiment is concerned with the quantification of this effect.

Purpose. To determine the amount by which a circle must be rotated before it appears as an ellipse of a given shape.

Materials. A circle is mounted on an upright rod so that the disk is in the vertical plane. In addition, several ellipses differing in width but all of them the same height as the circle are also mounted on vertical rods. These ellipses are mounted so that the long (major) axis is vertical. Pro-

vision is made by means of a pointer and protractor to determine the amount by which the rod and circle have been rotated. Instead of being calibrated in degrees, the protractor should give the projection of the diameter of the circle on a horizontal line in the plane of the ellipse. The protractor may be calibrated by means of the following formula

$$P = d \cos \Theta$$

Where P denotes the projection, d is the diameter of the circle, and Θ is the angle of rotation.

Procedure. The circle and an ellipse are placed perpendicular to the line of the subject's sight at a distance of about 10 feet. The experimenter then rotates the circle until the subject reports that it matches the ellipse in shape. At that point, he reads off from the protractor the value of the projection of the diameter of the circle. On one half of the trials, the experimenter begins with the circle perpendicular to the line of sight; on the other half of the trials, the circle is initially parallel to the line of sight. There should be about ten trials divided equally between the two starting positions. The starting positions should be alternated in random order. A new ellipse may then be substituted and the entire series repeated.

Treatment of Results. The larger the angle of rotation, the smaller the projection of the diameter of the circle needed to match the ellipse, and hence the greater the amount of constancy. The amount of constancy may be conveniently quantified by comparing the projection with the width (minor axis) of the ellipse. For this purpose, the average of the measurements of projection is computed. The ratio, $\frac{\text{Width of Ellipse}}{\text{Mean Projection}}$, provides us with an index of constancy. If there were no constancy effect, the width of the ellipse and the mean projection would be equal, and the ratio would be 1. The greater the constancy, the smaller the projection, and hence the greater the ratio.

General References

Boring, E. G., *Sensation and perception in the history of experimental psychology,* New York: D. Appleton-Century Co., 1942, Chap. 7.

Woodworth, R. S., *Experimental psychology,* New York: Henry Holt and Co., 1938, Chap. 25.

Figure and Ground

Hartmann, G. W., *Gestalt psychology,* New York: The Ronald Press, 1935, Chap. 6.

Koffka, K., *Principles of Gestalt psychology*, New York: Harcourt, Brace and Co., 1935, Chap. 5.

Rubin, E., *Visuell wahrgenomnene Figuren*, Copenhagen: Glydendalske Boghandel, 1921.

Wever, E. G., Figure and ground in the visual perception of form, *Amer. J. Psychol.*, 1927, 38:194-226.

Perceptual Grouping

Hartmann, G. W., *Gestalt psychology*, New York: The Ronald Press, 1935, Chap. 6.

Wertheimer, M., Untersuchungen zur Lehre von der Gestalt, *Psychol. Forsch.*, 1923, *4*:301-350. English translation in W. D. Ellis (ed.), *A source book of Gestalt psychology*, New York: Harcourt, Brace and Co., 1938.

Form Constancy

Koffka, K., *Principles of Gestalt psychology*, New York: Harcourt, Brace and Co., 1935, Chap. 6.

Sheehan, M. R., A study of individual consistency in phenomenal constancy, *Arch. Psychol.*, 1938, No. 222.

Thouless, R. H., Phenomenal regression to the real object, *Brit. J. Psychol.*, 1931, *21*:339-359; 1931, 22:1-30.

Geometrical Illusions

Boring, E. G., *Sensation and perception in the history of experimental psychology*, New York: D. Appleton-Century Co., 1942, pp. 238-245.

»»» 9 «««

PERCEPTION OF SPACE

THE objects that we see, hear, touch, and manipulate are localized in space: they occupy a definite position with reference to our bodies. Among these objects are parts of our own bodies as well as "things out there." We take the localization of objects in space for granted and, indeed, find it hard to imagine any world but a spatial world. When we stop to consider how living organisms manage to behave adequately in space, to move about in three dimensions, and to localize objects with astonishing precision, we soon find that the perception of space is highly complicated and must call into operation the utmost in constructive and integrative activity on the part of the perceiver.

SENSORY SYSTEMS IN SPACE PERCEPTION

Man relies heavily upon his eyes in his spatial adjustments, and in our thinking we tend to equate the spatial world with the visual world. Vitally important as is the visual system in the construction of our spatial world, it is by no means the only source of our knowledge about space. In addition to the visual system, there are two other important sensory systems which provide the perceiver with information about the location of objects and events in space. We not only see things in definite location, we also hear them as coming from the right or left, from above or below, and we, moreover, feel them as touching our hand or arm or some other part of our bodies.

When we need to localize objects in space, and in particular parts of our own bodies, we are further aided by sensory systems which are peculiarly adapted to this function. First of all, there is the kinesthetic system. Our muscles, tendons, and joints are equipped with receptors which respond to movements of these parts. The

"proprioceptive impulses" coming from the muscles are particularly important in the maintenance of posture. In short, the kinesthetic system tells us, as it were, about the movements of our bodies in space. Aiding this system is the vestibular system. The semicircular canals and the vestibule of the inner ear contain receptors highly sensitive to movement and acceleration. To summarize, then, several major sensory systems participate in the adjustments which man makes to the location of objects and himself in space. Certainly, in the location of objects in the environment, the visual system is primary, even though it is supplemented continually by the auditory and the somesthetic systems. The great precision and success of our spatial orientation are due in no small measure to the coöperation of these three systems. As we move about in our environment, intersensory coöperation is continually checked against the results of our movements and our manipulation of the objects. In this chapter we will limit ourselves to a discussion of the contribution which the visual system makes to the perception of space.

Basic Visual Conditions

The visual perception of space, as we know it, would be impossible if it were not for certain functional properties of the optical process. These properties insure the stability of our visual world and the continuing success and precision of spatial adjustments. Objects fixed in space usually appear so and remain so, and the spatial relations among objects show a similar stability. Let us briefly consider the stabilizing properties of the optical process.

Correspondence Between Retinal Image and Physical Object. It is probably a truism by now to say that we do not "see" our retinal images, that the formation of the retinal image is only one of the long chain of events necessary for the perception of an object. Nevertheless, if the retinal image did not systematically represent certain features of the physical object, a stable visual world would be impossible. We are referring here to the eye as an optical instrument which forms a retinal image preserving the spatial relations characteristic of the physical object. The outside edges of the physical object are represented on the outside edges of the retinal image. The relations among the parts of the object remain basically unchanged in their representation on the retina.

The fact that the retinal image is inverted with respect to the physical object is irrelevant. The inversion of this image with respect to the object is a physical fact that does not present a special problem in the analysis of perception. We must say again that we do not "see" the retinal image. We learn that the stimulation of the bottoms of retinae means "up." The relations among the parts of the physical object represented in the image are still invariant.

Law of the Visual Angle. The stability of the visual world is further aided by another optical property of the eye. As an object recedes from the eye, the size of the retinal image decreases in accordance with the *law of the visual angle*. This law states that the linear size of the optical image is inversely proportional to the distance of the object. Thus, if we double the distance of an object from the eye, we halve the height and width of the retinal image. This law does not imply that we see things in exact accordance with it. A man 20 feet away does not appear half as tall as he does at 10 feet. As we will find in our discussion of size constancy, there may be appreciable variation in the distance of an object from the eye without corresponding changes in perceived size. The important point is that there is a lawful relationship between distance and size of retinal image. Unless this were the case, any lawful relationship between distance and perceived size would be impossible.

Resolving Power of the Visual System. When the eye focuses upon two objects in the environment, two retinal images are formed, and the visual system maintains this resolution to an astonishing degree. We have already discussed the nature and conditions of *visual acuity* in Chapter 6. Unless two objects are seen, we could not perceive any relation between them. The visual perception of space is based to a considerable extent upon the spatial relations among objects. Therefore, we must first resolve these objects before their relations can become effective in perception.

Thus, the optical properties of the eye and the resolving power of the visual system provide basic conditions for the visual perception of space. With this background, we can now proceed to an examination of the specific determinants of space perception.

The Spatial Framework

Our visual world is tridimensional. Objects are localized not only

up and down, or to the right and left, but also near and far. Furthermore, solid objects are seen as solid, i.e., in depth. There is, then, a spatial framework which must be described in terms of three dimensions. An important property of this framework is that it is anchored by two main directions: the vertical and the horizontal. These two directions exhibit a high degree of stability and provide a framework for the localization of objects.

What are the conditions for the establishment of this framework? How do these stable directions come into being? We may first note that there is no one-to-one correspondence between the direction of the lines of the retinal image and the perceived direction of those lines. A pencil held upright casts a vertical line on the retina and is seen as upright. If the observer now tilts his head, the retinal image of the pencil will be oblique with respect to its former position, yet it will still appear as vertical. Moreover, objects casting retinal lines that have the same direction may not so appear. Thus, when the head is in a certain position, a ruler on the desk may cast a vertical retinal line and so may a book standing on the shelf above the desk. The ruler will appear horizontal and the book, vertical. Thus; the direction of the retinal line does not in itself determine the perceived direction.

What is seen as vertical and what is seen as horizontal depends upon the relations among the parts of the visual field. In a given visual field, there are usually *main lines of organization*. The sides of a table, the windows and doors of a room, are examples of objects providing main lines of organization, i.e., lines that serve as frames for a diversity of objects. It is these main lines of organization which tend to take on the horizontal and vertical directions. The directions of other lines in the visual field depend upon their relation to the main lines of organization.

The organization of the spatial framework is flexible, as the following demonstration clearly shows. A mirror is tilted in such a way that lines habitually seen as vertical or horizontal are seen as oblique by an observer looking at the mirror through a tube. The tube is arranged so that it limits the subject's view to the mirror. Initially, the field as seen in the mirror is "out of joint." Lines that should look vertical are seen as oblique. As the observer continues

to view the field in the mirror, the main lines of organization "right themselves" and appear as horizontal or vertical.

Within this spatial framework, anchored by the horizontal and vertical directions, we learn to move about and to manipulate objects in the environment. The various sensory systems coöperate closely in our spatial adjustments. If we snap our fingers, we localize them by seeing, hearing, and feeling them in the same spatial position. This intersensory coöperation is modifiable within limits. An enterprising psychologist inverted his visual field by wearing a system of lenses. Objects which we ordinarily locate at our right appeared on the left, and objects ordinarily seen above us appeared below. This inversion of the visual field created serious intersensory discrepancies. Thus, the investigator would see a clock in one place and hear it ticking in another. Similarly, he would feel his feet below him, and they would appear above his head. To a certain extent, it was possible to adjust successfully to this confusing situation by ignoring visual cues. Much more striking is the fact that an intersensory reëducation took place. As he continued to wear the lenses, the intersensory discrepancies became less and less noticeable. He saw objects where he heard and where he felt them. However, he continued to call objects "up" which were normally "down" and "left" those which we observers without inverting lenses call "right." In short, the "naming" of the direction did not change, but the behavioral meaning was changed in accordance with the requirements of successful locomotion and manipulation. Indeed, upon removal of the lenses, a brief period of readjustment was necessary.

The Perception of Distance

The images cast on the retina are two-dimensional and they preserve the correct spatial relations among parts of the visual field. The identification of "right and left" and "up and down" are matters of relative judgment made possible by this correct preservation of the spatial relationships. But we see objects as far and near, solid as well as flat. The question at once arises as to how two-dimensional retinal images lead to the perception of space in depth. It seems as if the perceiver uses certain clues and indications provided by these two-dimensional images to infer or construct a tri-

dimensional space. We must emphasize the phrase "as if." There is no reason to believe that any process akin to logical inference takes place. Tridimensionality is immediately given.

Monocular Determinants of Perceived Distance

Let us see if *one eye* can offer any clues to the perception of distance. By having the observer close one eye, we limit the number of clues available and separate them from those that depend upon the simultaneous functioning of both eyes. Let us list and briefly discuss monocular clues to distance. Most of these determinants are rather obvious and are safely inferred from their effectiveness in daily experience and especially in the visual arts.

Apparent Size of Familiar Objects. Automobiles viewed from the tower of the Empire State Building look more like oversized ants than they do like cars. We see them, however, as distant cars rather than oversized ants at close range. What the observer does in effect is to utilize the law of the visual angle. As a car recedes from the eye, the retinal image grows smaller and smaller. Clearly, this clue can be effective only with familiar objects. The implausible stranger who had never seen a car could not judge its distance by its apparent size alone. An especially striking instance of such *linear perspective* is the familiar convergence of parallel railroad tracks in the distance. The distance between the tracks is, of course, constant, but the corresponding retinal image, and hence the apparent separation of the tracks, decreases in size the farther away the tracks are. The artist can make good use of linear perspective to create the impression of distance on a two-dimensional canvas.[1]

Interposition of Objects. If our view of one object is obstructed by the presence of another, we will see the obstructed one as farther away. If a tree covers part of our view of a house, the tree appears nearer than the house. Conversely, if the house obstructs our view of the tree, we see the house nearer than the tree.

Clearness of Detail. The perception of details is a matter of visual acuity. The detail of a brick wall becomes lost in the distance because the visual angle subtended by the individual bricks de-

[1] This and other illustrations are not offered as experimental proof of the importance of the determinant in question. Obviously, in these illustrations, the effects of other determinants are not held constant as distance varies. The examples are given as aids in understanding the nature of the determinant. Only under controlled conditions can conclusive proof be achieved.

creases with distance. Thus, when we cannot see the details of the wall, we see it as far away. When each brick stands out clearly, we see it as nearby.

Changes in Color. The color of an object may undergo changes as the light waves reflected from the object travel through the haze of the atmosphere *(aerial perspective)*. Not only does the color tend to be more bluish as distance is increased, but the apparent brightness of the object diminishes as well. The green leaves of a distant tree take on a bluish tinge. The bright, black car shines less as it speeds away.

Lights and Shadows. If we know the direction of the illumination, then an object standing in a shadow is seen to bear a definite spatial relation to the object casting the shadow. The pattern of a shadow and its location help us to judge whether there is a depression or an elevation in the land. If we are looking toward the sun and the shadow is on the far side (near the sun) of an irregularity in the ground, then we see a depression. On the other hand, if the shadow is on the near side (near us), then we see an elevation.

Movement Parallax. Perception of relative movement aids in judging distance. Let us illustrate this determinant by the following demonstration. Close one eye and hold up two fingers, one behind the other and about 10 inches apart. Fixate the far finger. Now move the head from side to side, and you will see the near finger moving in the direction opposite to the movements of the head. This phenomenon is known as *movement parallax*. It is sometimes referred to as *monocular parallax,* because it can be obtained with the use of only one eye.

Let us consider the explanation of the movement of the near finger while the far finger is fixated. As the head is moved from side to side, the image of the far finger remains stationary on the retina by virtue of the fixation. When the head is moved to the right, the retinal image of the near finger also moves to the right. If we were to accomplish this retinal displacement of the near finger by moving *it* instead of the head, this finger would have to move to the left. It, therefore, appears to do so.

Accommodation. All of the determinants discussed above are visual in nature, i.e., they are characteristics of the visual field. There is the possibility, however, that proprioceptive impulses furnish auxilliary clues in perceiving the distance of near objects

These impulses come from the muscles that control the accommodation of the lens to varying distances. It is through this accommodation process that clear vision is achieved at different distances. Many ingenious experiments have been designed to demonstrate the effectiveness of accommodation as a clue for perceived distance. Correlations between changes in accommodation and perceived distance have been reported. These results do not provide us with

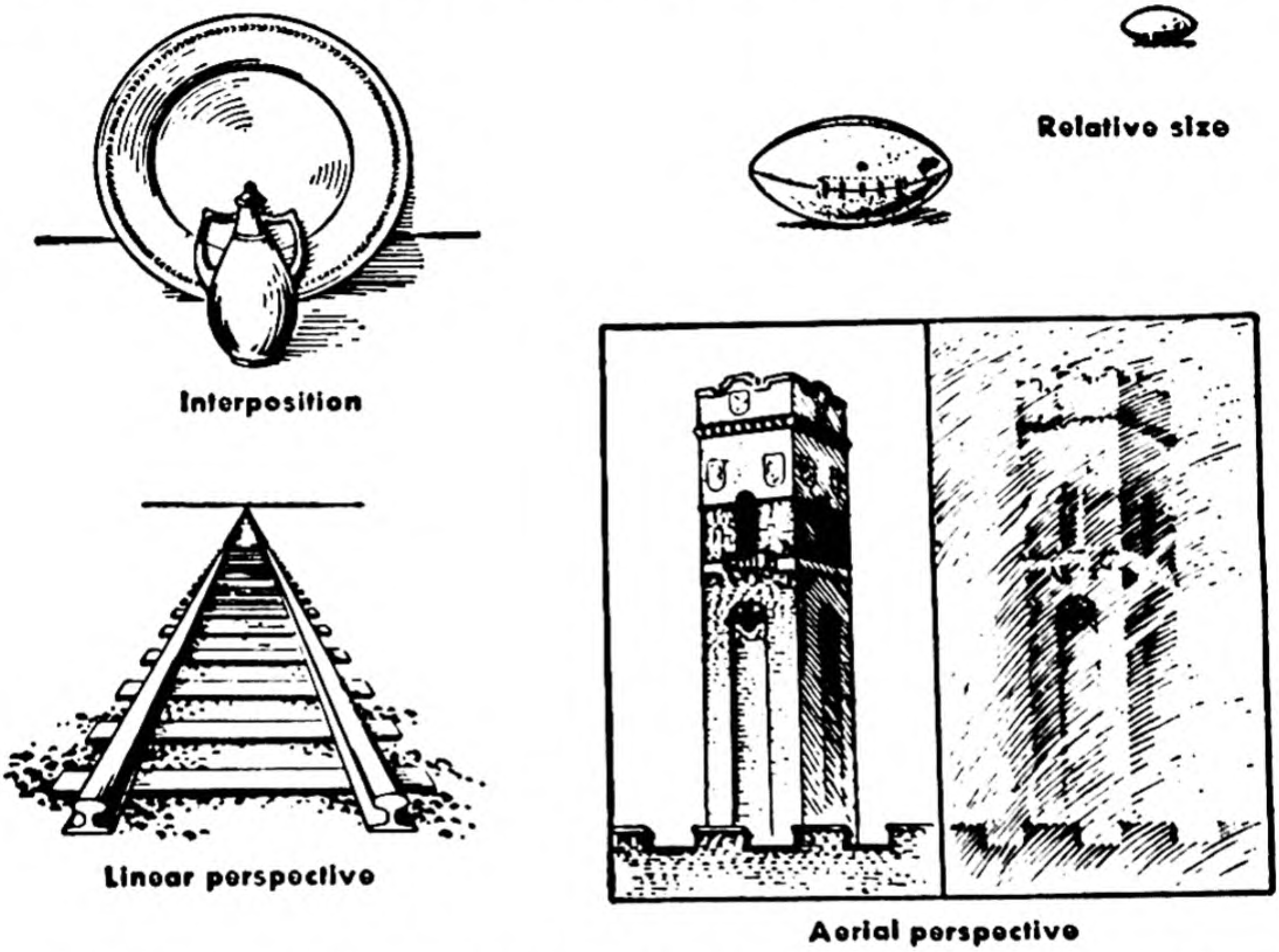


Fig. 55. Some Monocular Determinants of Perceived Distance. (From N. L. Munn, *Psychology*, 1946, p. 154, by permission of Houghton Mifflin Company.)

conclusive evidence, however, concerning the effectiveness of nonvisual, proprioceptive clues to distance. For one thing, it is difficult indeed to hold all factors other than accommodation constant, especially since accommodation and convergence tend to work together. Even if accommodation is varied independently of convergence, changes in accommodation must change the characteristics of the visual field. Changes in accommodation affect the clearness with

which details are seen, that is, the amount of microstructure perceived. These changes in microstructure provide visual clues for distance. Even if the nerve impulses arising from the accommoda-

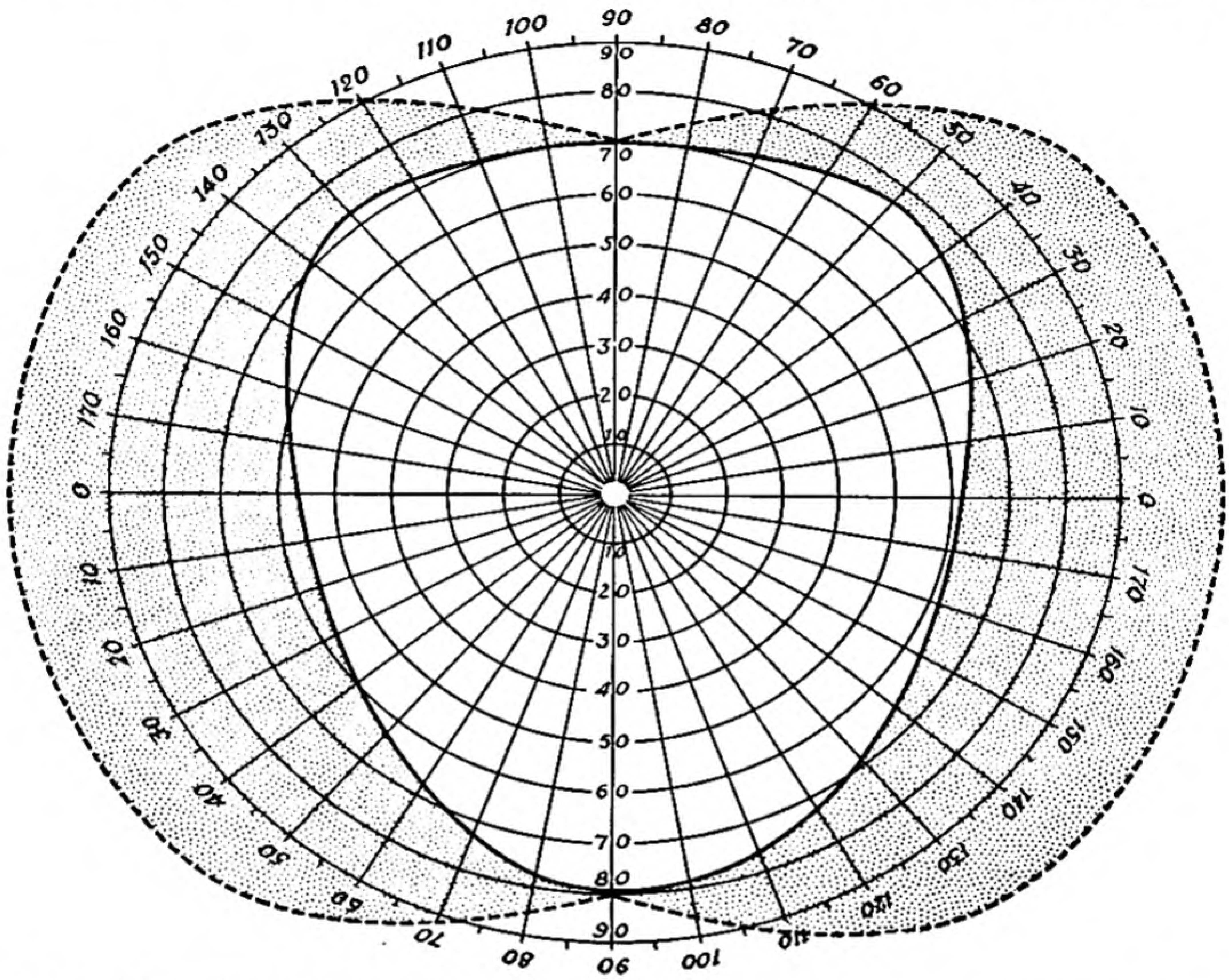


FIG. 56. The Field of Binocular Vision. The white area represents the field of overlap, i.e., the part of the field seen by both eyes when fixation is at the center of the field. The two shaded areas represent the nonoverlapping fields. The numbers along the vertical meridian refer to distance from the fovea in degrees. The numbers around the periphery designate the angle of elevation. (From *Introduction to physiological optics* by J. P. C. Southall, p. 209. Copyright 1937 by Oxford University Press, Inc.)

tion process do contribute information about distance, they certainly play a minor role relative to the visual determinants.

Binocular Determinants of Perceived Distance

Some of the most effective determinants of visual space perception stem from the fact that we have two eyes located about 2.5

inches apart. To understand the nature of these binocular determinants, let us briefly review the main facts of binocular vision.

Field of Binocular Vision. Each of the two eyes commands a visual field of approximately 130 degrees. The extent of the *monocular* field may be determined by closing each eye in turn. The visual field is then determined by the extent of the environment that can be seen under steady fixation. The monocular field is not homogeneous with respect to various visual characteristics. Most important among these are differences in amount and sharpness of detail. Maximum detail in clearness is achieved when the

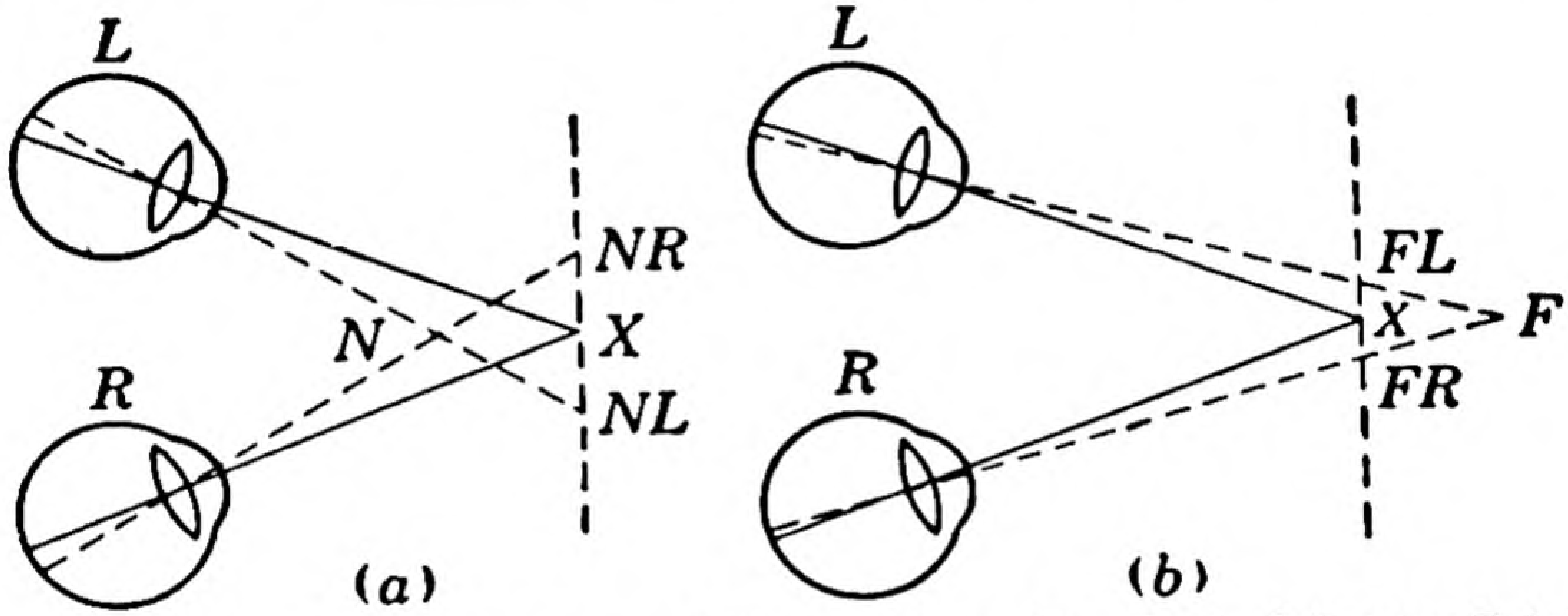


FIG. 57. Corresponding and Noncorresponding Retinal Points. Point *X* is the point of fixation and is seen as single. Points *N* and *F* can be seen as double. The images of Point *N* are crossed, whereas the images of Point *F* are uncrossed. (Reproduced by permission from *Foundations of Psychology* by Boring, Langfeld, and Weld, p. 135, published by John Wiley & Sons, Inc., 1948.)

retinal image of an object is cast upon the fovea. The line connecting the retinal image on the fovea and the corresponding object is called the *line of regard.*

When both eyes are open, there results an extensive field which is common to both eyes. This area of overlap is the *binocular field of vision.* Normally, the eyes converge upon a given object, i.e., the two lines of regard cross at the point of fixation. If an object on the horizon is viewed, the lines of regard are parallel, i.e., there is no convergence. Whether or not the eyes are converged, the field seen in common is the binocular field. A graphic picture of the monocular fields of the two eyes and the binocular overlap is shown in Fig. 56.

Corresponding and Noncorresponding Retinal Points. When the eyes are converged upon a point of fixation, this point is seen as single. In Fig. 57, for example, point X is the point of fixation and is seen as single. We call the points on the retinae of the two eyes that are stimulated by the single point X, *corresponding retinal points.* In the same figure, the point N on the near side of X

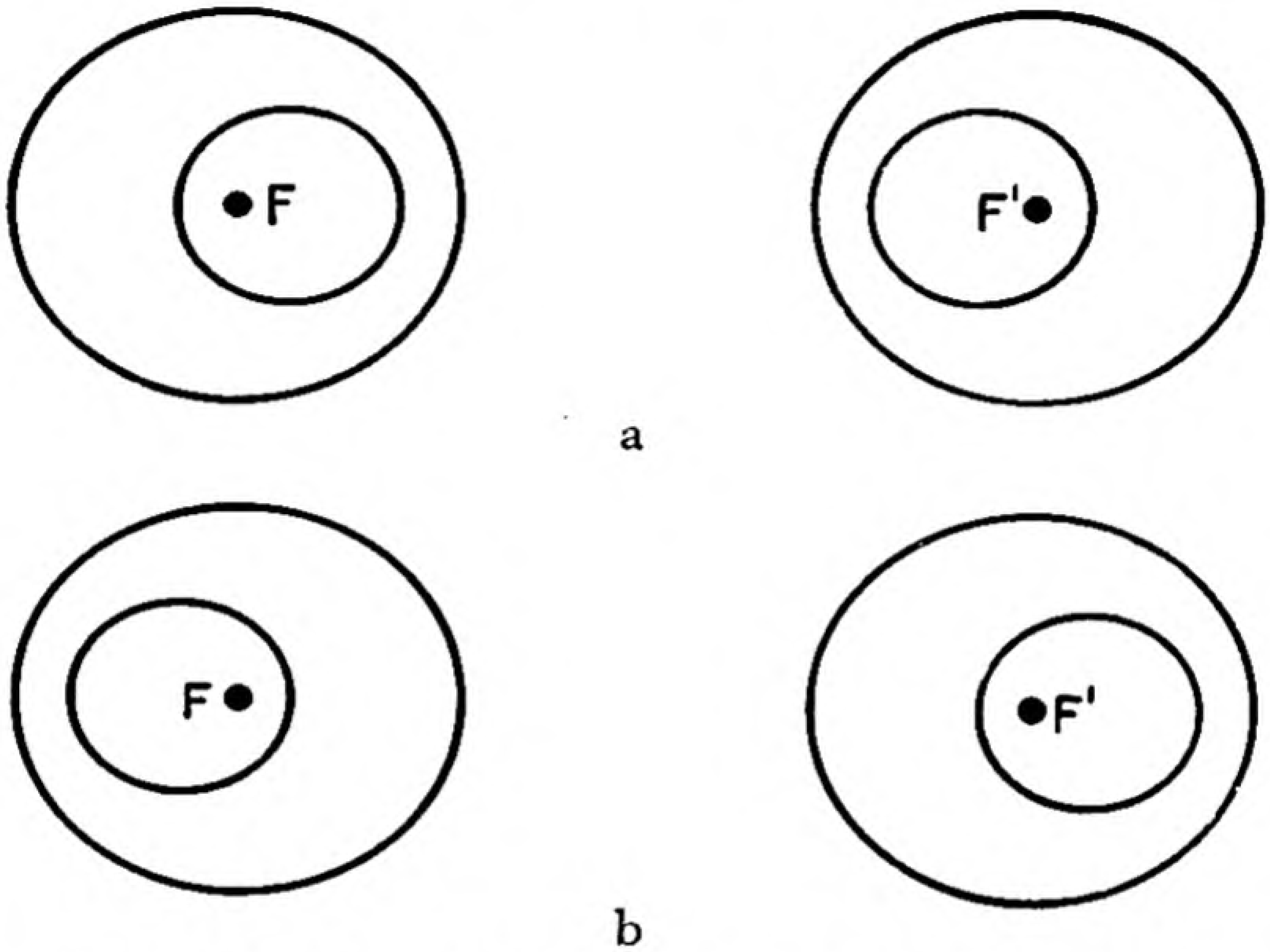


FIG. 58. How the Two Eyes See a Truncated Cone, with the Axis of the Cone Horizontal and Pointing at the Observer. F and F' represent the fixation point. In (a) the small end of the cone is near the eye; in (b) the large end is near the eye.

can be seen as double. Similarly, point F, on the far side of X, can also be seen as double. When a single point can be seen as double, we call the two retinal points stimulated by it, one on each retina, *noncorresponding points.* It should be noted that corresponding points are defined in terms of singleness of vision and not in terms of anatomical location. By mapping the locus of single vision (*horopter*), it can be shown, however, that corresponding points are symmetrically located with respect to the foveae. Thus, if the two retinae were superposed, corresponding points would tend to fall on top of each other.

Binocular Parallax. When the eyes view a solid object, rather than a single point, they necessarily obtain somewhat different views of that object. This disparity in the two retinal images is known as *binocular parallax.* Recall the fact that the eyes are about 2.5 inches apart. When they fixate a single point on a solid object, the horizontal pattern of the retinal images must be somewhat different. Thus, the right eye will see a little more of the right side

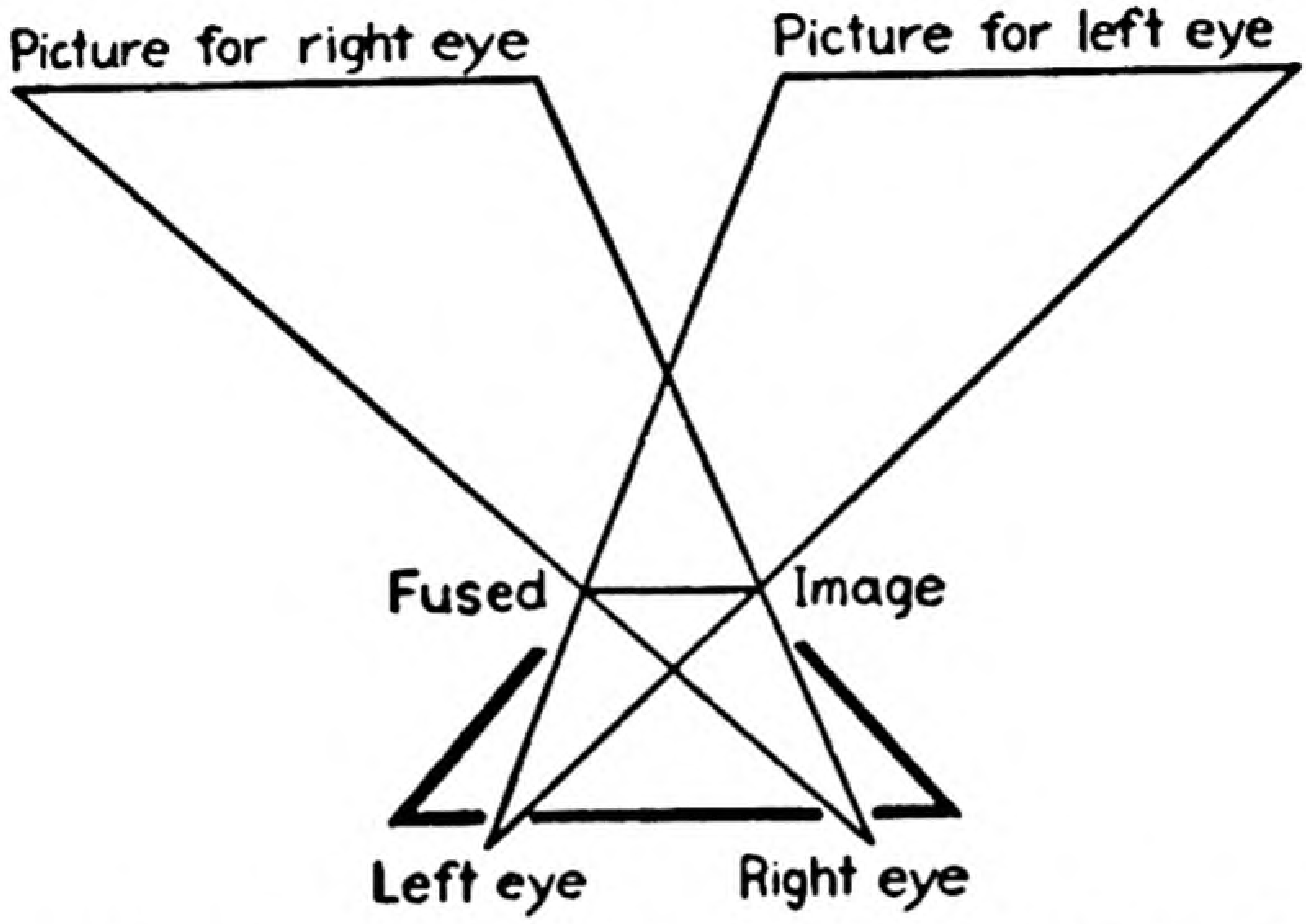


FIG. 59. A Simple Box Stereoscope. (From *Introduction to physiological optics* by J. P. C. Southall, p. 238. Copyright 1937 by Oxford University Press, Inc.)

of the object, and the left eye, a little more of the left side. Suppose, for example, a truncated cone is viewed with the axis of the cone horizontal and pointing at the observer. If the small end of the cone is near the eyes and the large end away, the retinal images for the two eyes will be as shown in Fig. 58a. It should be clear that the left-hand figure represents the retinal image for the left eye and the right-hand figure, the image on the right retina. Fig. 58b shows what happens to the retinal images when the large end of the cone is placed near to the eyes. Examination of these retinal patterns shows that the large circles fall on corresponding points of the retinae. Since there is a disparity in the two images, the small circles do not fall on corresponding points. It is in the nature of binocular

parallax that parts of the images fall on corresponding points whereas other parts do not.

The two images being disparate, we might expect double images to be seen. In actual fact, however, the observer does not see two small overlapping circles enclosed by one large one. Rather, he sees a truncated cone in depth. The fusion of the disparate retinal

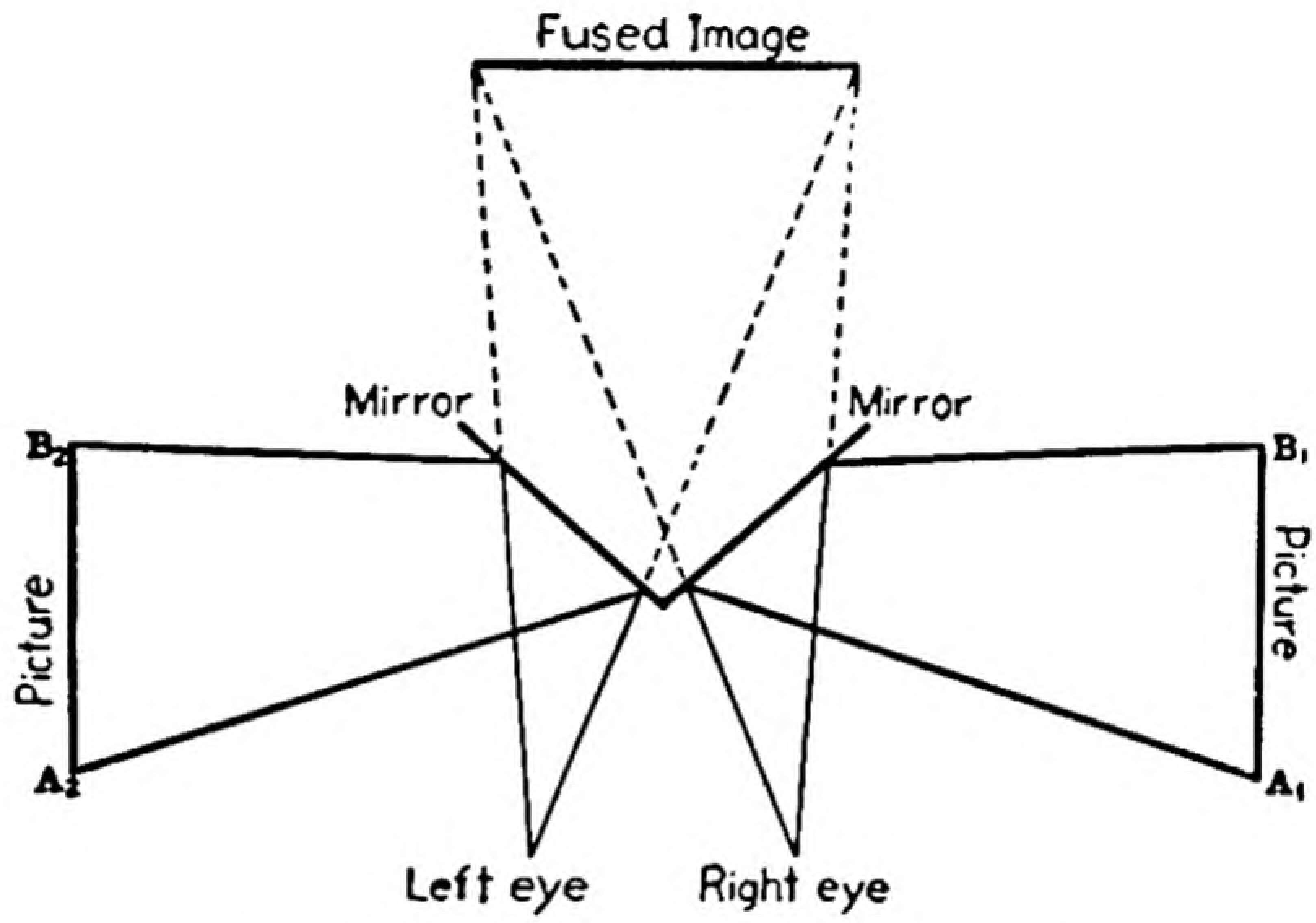


FIG. 60. A Mirror Stereoscope. (From *Introduction to physiological optics* by J. P. C. Southall, p. 240. Copyright 1937 by Oxford University Press, Inc.)

images produces the depth effect. Binocular parallax is, indeed, one of the most important determinants of perceived depth.

Stereoscopic Vision. One might expect that a depth effect could be produced by the use of two simple line drawings, such as those shown in Fig. 58, one presented to each eye. All that is necessary is an instrument to present the drawings separately to the two eyes. Such an optical instrument is known as a *stereoscope.* The basic principle of the stereoscope is illustrated in Fig. 59. This simple box stereoscope merely insures that the picture for the left eye does not stimulate the right eye, nor that the picture for the right stimulates the left eye. This simple box stereoscope has certain disadvantages. Convergence, for example, is rather difficult to obtain. More

elaborate and flexible stereoscopes have been devised for producing depth effects.

1. *The mirror stereoscope.* (This stereoscope is also known as the Wheatstone stereoscope, after the inventor.) Fig. 60 shows how two stereoscopic pictures can be presented for fusion. The mirrors provide that the pictures are separately presented, one to each

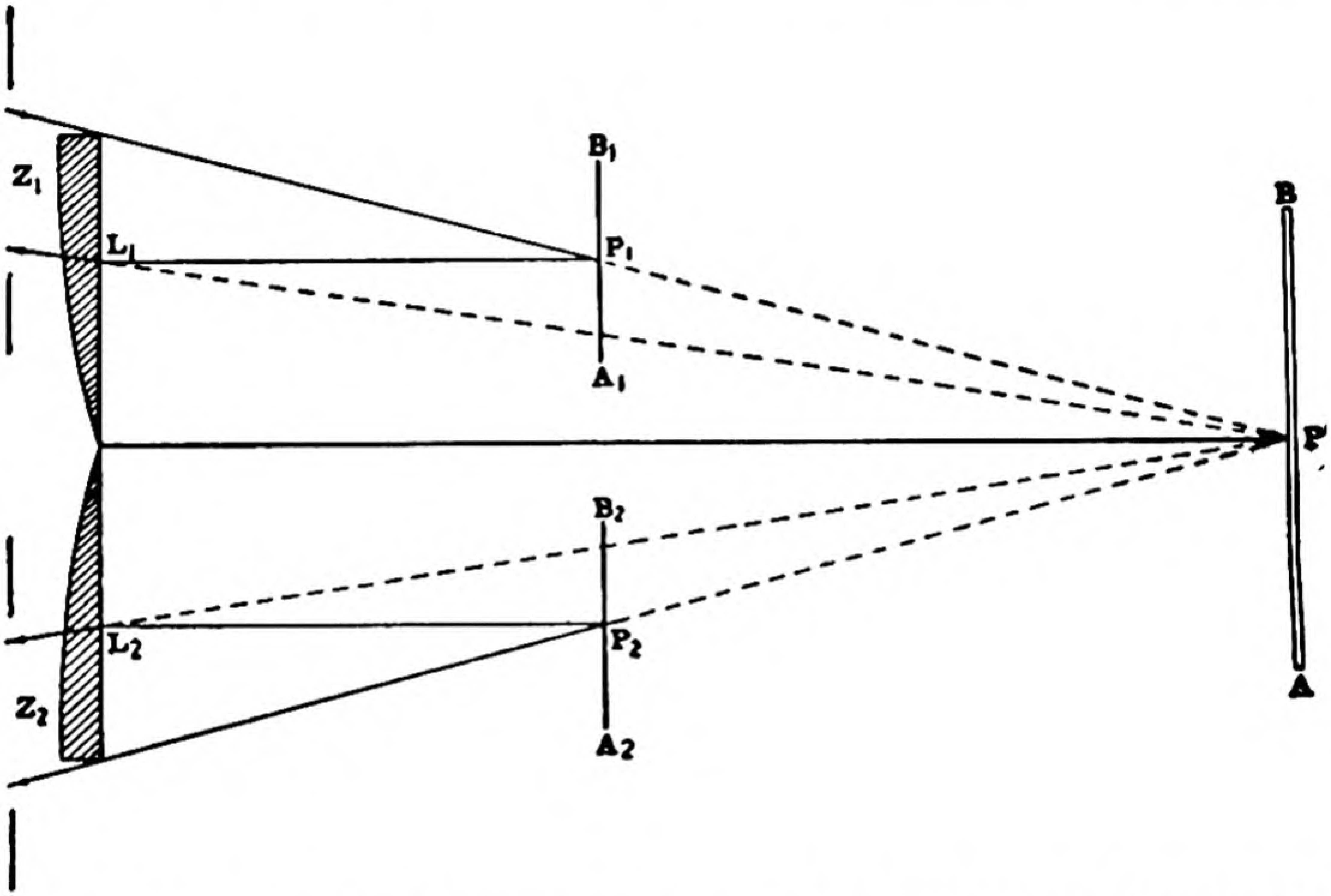


FIG. 61. A Prism Stereoscope. (From *Introduction to physiological optics* by J. P. C. Southall, p. 242. Copyright 1937 by Oxford University Press, Inc.)

eye. The angle between the mirrors is adjusted to give the desired degree of convergence. By altering the distance between the pictures and the mirrors, the proper degree of accommodation can be obtained. Note the apparent location of the fused image.

2. *The prism stereoscope.* (This stereoscope is also known as the Brewster stereoscope, after the inventor.) It is also possible to present the pictures separately by the use of prisms. The prisms refract the light waves, changing their direction so as to permit

the eyes to converge. Fig. 61 shows one such prism stereoscope. L_1 and L_2 are the two prismatic lenses. The bases of these prisms are placed on the outside so that the light rays coming from P_1 or P_2 are refracted, thus causing the eyes to converge on the point P. The pictures, or stereograms, are represented by A_1B_1 and A_2B_2, respectively, while AB represents the apparent location of the fused picture seen in depth.

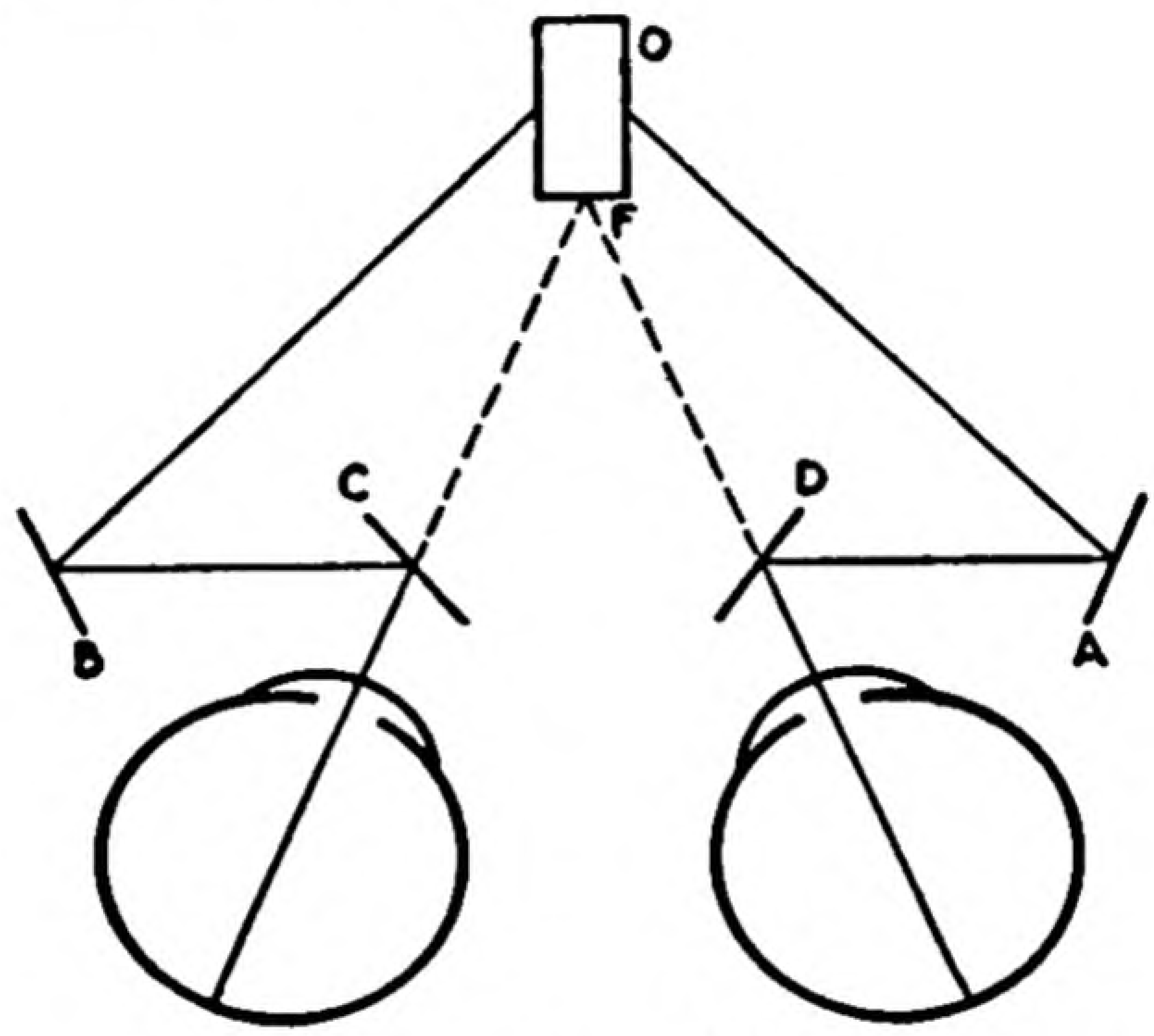


FIG. 62. A Telestereoscope. (From H. A. Carr, *An introduction to space perception*, 1935, p. 214, by permission of Longmans, Green and Co., Inc.)

3. *The telestereoscope*. For certain purposes, it is advantageous to exaggerate the degree of retinal disparity. This increased disparity can be accomplished by the use of a telestereoscope. The basic principle of this device is illustrated in Fig. 62. It is obvious that the effective separation between the two viewing points—mirrors A and B—is greater than the interocular distance. The rays reflected from mirrors A and B are sent to the eyes via mirrors C and D, respectively. By increasing the distance between A and B, the degree of retinal disparity can be made much greater than normal. Consequently, the apparent depth is greatly enhanced. Thus, even if two objects are far away, it may be pos-

sible by the use of this instrument to tell which is the farther. This principle is employed in many range-finding devices.

4. *The pseudoscope.* This instrument makes it possible to present to the left eye the image normally falling on the right eye, and *vice versa.* In this manner the retinal disparity of the views is inverted. It is possible to invert the retinal images either horizontally, vertically, or both. It is by means of a pseudoscope that the studies on the effects of retinal inversion cited on p. 181 were conducted.

Uses of the Stereoscope. Pictures for stereoscopic presentation may be either drawings or photographs. If the two pictures are identical, no depth effect is obtained. When the two pictures are disparate, stereoscopic fusion in depth may occur. The degree of depth depends upon the amount of disparity. Striking effects may be obtained with the aid of a camera, taking two pictures with the camera successively placed at positions separated by the interocular distance or more. The simplest demonstrations for experimental purposes are obtained by line drawings such as shown in Fig. 58. Stereoscopy has been put to a variety of uses.

Demonstration of binocular fusion. Stereoscopes are a favorite laboratory device for the demonstration of binocular parallax as a determinant of perceived depth.

Studying the effect of convergence and accommodation. As we have seen, both mirrors and prisms may be used to change the degree of convergence of the eyes when viewing stereograms. Furthermore, by altering the distance between the stereogram and the reflecting mirror, the degree of accommodation is influenced. Special types of stereoscopes have been devised in an attempt to vary convergence and accommodation independently. Studies concerned with the effects of accommodation and convergence on perceived distance have frequently employed stereoscopic presentation of stimulus materials.

In summary, then, the stereoscope derives its importance from the fact that it serves to control and measure the amount of retinal disparity which is so dramatically effective in the perception of depth.

Diplopia. Stereoscopic fusion, important as it is, is not the only binocular determinant of perceived depth. Double images

serve as effective binocular clues. Let us return to Fig. 57 on p. 186. When the eyes are converged on *X*, points nearer or farther away from the eyes than *X* stimulate noncorresponding points of the retinae. As a result, such points as *N* and *F* can be seen as double. Although both *N* and *F* can be seen as double, there is an important difference between the near point *N* and the far point *F*. The two images of *N* are *crossed*, i.e., the left eye sees the image on the right while the right eye sees the image on the left. In the case of *F*, however, the images are *uncrossed*, and each eye sees the image on its own side. These facts can be easily ascertained by again holding up two fingers, one behind the other. Now fixate the far finger. The near finger will be seen as double. That the images are crossed may be verified by closing first one eye and then the other. Similarly, fixate the near finger and obtain a double image of the far one. That the images are now uncrossed may be verified in the same manner as before.

It is clear from the diagrams of Fig. 57 that crossed versus uncrossed double images should be an important clue to distance. The point *N*, which produces crossed images, casts an image which on each retina is on the temporal side of the fovea. The point *F*, on the other hand, stimulates points nasally to the fovea. In short, if the double images are crossed, the object is nearer than the point of fixation. If the images are uncrossed, the object must be farther away than the point of fixation.

It may be objected that double images are frequently not seen when we would expect them to be, in terms of the geometry of the situation. Nor is it possible for the observer to tell crossed images from uncrossed ones without some explicit trial and check. How, then, can double images serve as clues to distance? We must reiterate that the determinants of a perceptual judgment need not be reportable in order to be effective. Some determinants act *as if* they were clues from which inferences are drawn, even though the process of inference, whatever its nature, cannot be verbalized. Whether or not a determinant is effective can be ascertained only on the basis of the differences which it causes in the subject's behavior. By this criterion, double images qualify as one of the determinants of perceived distance.

Seeing Behind Small Objects. There is another important

determinant of distance dependent upon binocular parallax. Fig. 63 shows a schematic diagram which is known as *Leonardo's paradox,* since it was first pointed out by Leonardo da Vinci. As this diagram shows, part *PG* of the background *PE* is obscured for the left eye by the object *C*. Similarly, *DE* cannot be seen by the right eye. What one eye cannot see, however, the other can, and the total background *PE* can be seen binocularly. This fact troubled da Vinci since he could not reproduce it on his canvas.

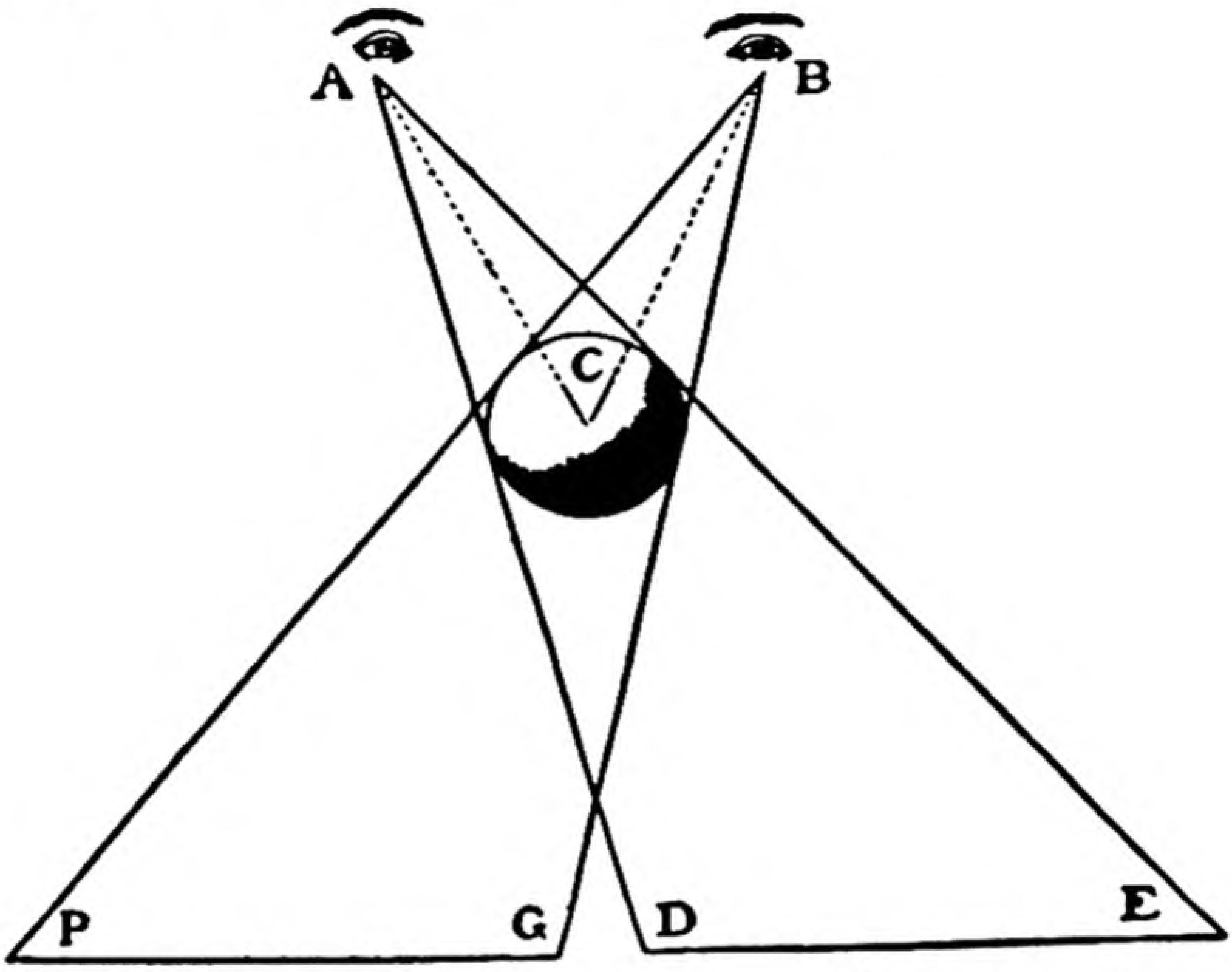


FIG. 63. Leonardo's Paradox. Although each eye can see only part of the background of the object *C*, the total background can be seen binocularly. (From E. G. Boring, *Sensation and perception in the history of experimental psychology,* 1942, p. 283, by permission of Appleton-Century-Crofts, Inc.)

Seeing behind small, near objects is a powerful clue to distance. The farther away an object is, the less it is possible to see fully behind it, and the more it obscures another object placed a constant distance beyond it. The relation is an exact one and, therefore, serves as a reliable determinant.

Convergence. As in the case of monocular vision, we conclude our discussion of binocular determinants with a nonvisual factor: convergence. The extrinsic muscles of the eye control convergence.

Their action may result in proprioceptive impulses which might act as clues to the distance of the object fixated. If such clues were effective, they would be limited to a range of about 60 feet from the eye, the range over which convergence occurs. Beyond this distance, the lines of regard are very nearly parallel. As in the case of accommodation, the experimental proof of the effectiveness of these proprioceptive impulses in the judgment of distance is by no means complete. The results of many ingenious experiments have remained contradictory. Some experimenters who have studied the effects of convergence in stereoscopic vision have reported that differences in convergence, with other factors held constant, significantly influenced the perception of distance. Whether or not they were indeed successful in eliminating or holding constant such visual factors as double images and, in some instances, retinal disparity, is still a moot question. Other experimenters have been convinced from their data that proprioceptive impulses from convergence and/or accommodation are ineffective. For example, when a bright circular spot, the visual angle of which is held constant, is moved back and forth in a long dark tunnel or hall, judgments of distance are either virtually impossible (monocular) or extremely poor (binocular). In this situation, such visual factors as retinal disparity are absent or minimized, and proprioceptive impulses aid the judgment of distance very little indeed, as compared with the normal effectiveness of visual clues. In both monocular and binocular vision, the predominance of visual determinants over nonvisual ones is well established.

The Perception of Size

Size Constancy. The problems of the perception of distance and of the perception of size are closely related. They are linked together by the law of the visual angle. This law would say that an object must become smaller and smaller as it recedes from the eye and that these decreases in size must be proportional to the distance. Since the object remains constant in physical size, the retinal image of it must continually decrease with distance. It is true that very distant objects look small. It is equally true that perceived size does not obey the law of the visual angle. Over a considerable range of distances, perceived size decreases much less

rapidly than would be predicted by the law of the visual angle. The failure of perceived size to decrease in proportion to the distance is known as *size constancy*. As in the cases of color and shape constancy, the term, constancy, is somewhat misleading. Perceived size neither varies with the visual angle nor does it remain invariant with distance.

Evidence for size constancy comes both from daily experience and from the laboratory. As an individual walks away from us on the street, he does not shrink in size. The chairs in a room look about the same size no matter where we are in the room. These examples could be multiplied, but we must turn to the laboratory for an analysis of the extent and determinants of size constancy.

Experimental Determination of Size Constancy. Perceived size can be specified and measured only with respect to some standard of reference. Suppose we present a subject with a circular disk 10 centimeters in diameter, and at a distance of 16 meters from the eyes. If we simply ask him about the size of the disk, even if he tells us the size in centimeters, let us say 12, we do not know what private yardstick he is using in making this judgment. Nor do we know whether the object looks to him like a 12-centimeter disk 16 meters away, or like a 12-centimeter disk one meter away. In short, perceived size is always a matter of relative judgment. Our task in measuring perceived size is to make the relativity explicit, to make the private yardstick a public one. It is for these reasons that in experiments on perceived size we require an observer to make a match or discriminate between a stimulus viewed under variable conditions and a stimulus viewed under standard conditions. Let us illustrate this procedure by describing a typical experiment on size constancy. We wish to find out how the perceived size of a circular disk varies as a function of its distance from the observer. This disk is 10 centimeters in diameter, and we first present it to the observer at a distance of 1 meter. The subject is provided with another disk, the diameter of which he can vary at will, and the distance of which remains fixed at 1 meter. The subject adjusts the diameter of the reference disk until it appears equal to the test disk. The test disk is then moved successively to distances of 2, 4, 8, and 16 meters, and at each distance the diameter of the reference disk is adjusted to a match. We can then plot the average

adjustment of the reference disk against the distance from the observer of the test disk.

Of course, the experiment could be conducted in a somewhat different way without altering the basic logic of the procedure. The diameter of the reference disk could be held constant, and the diameter of the test disk varied at each of the distances until a match was obtained. We would then plot against distance the diameter of the test disk required to match the reference disk.

Within this experimental framework, we can investigate perceived size as a function of such controllable variables as binocular versus monocular viewing, the environmental context, the degree of illumination, the subject's attitude, and so on. The aim is always to obtain a functional relationship in which perceived size is the dependent variable.

Perceived Size and Distance. When we plot perceived size against distance, we usually obtain a function lying somewhere between two limiting loci. At one extreme is a curve defined by the law of the visual angle. It is the line that would represent the data if perceived size varied inversely with the distance. At the other extreme is the horizontal line defined by perfect size constancy, i.e., perceived size as invariant with distance.[2] These theoretical lines can be found in Fig. 64. Actually, perceived size neither obeys the law of the visual angle nor is it invariant with distance. Where, between these extremes, the empirical function lies depends upon several parameters. Let us reëxamine Fig. 64. It depicts the results of an experiment in which the amount of environmental clues available to the subject was systematically varied. There were four experimental conditions representing a successive reduction of the environmental context. First of all, there was unimpeded binocular vision in which the subject could fully utilize all the relations present in his field of vision. As a first reduction in the number of available clues, the subject was restricted to the use of one eye. Further restrictions are represented by monocular vision with an artificial pupil and a reduction tunnel. This reduction tunnel, consisting of heavy black cloth, limited the view to the test

[2] Sometimes constancy may, as it were, overshoot the mark, and the perceived size increases for a time with distance. It is as if the observer overcompensated for increases in distance.

disk almost entirely. As Fig. 64 shows, the function relating perceived size to distance is systematically displaced as environmental context is varied. With unimpeded binocular regard, there is more than perfect constancy—the subjects overcompensate for distance. With monocular vision, there is complete size constancy, that is, perceived size is invariant with distance. Under the more restrictive

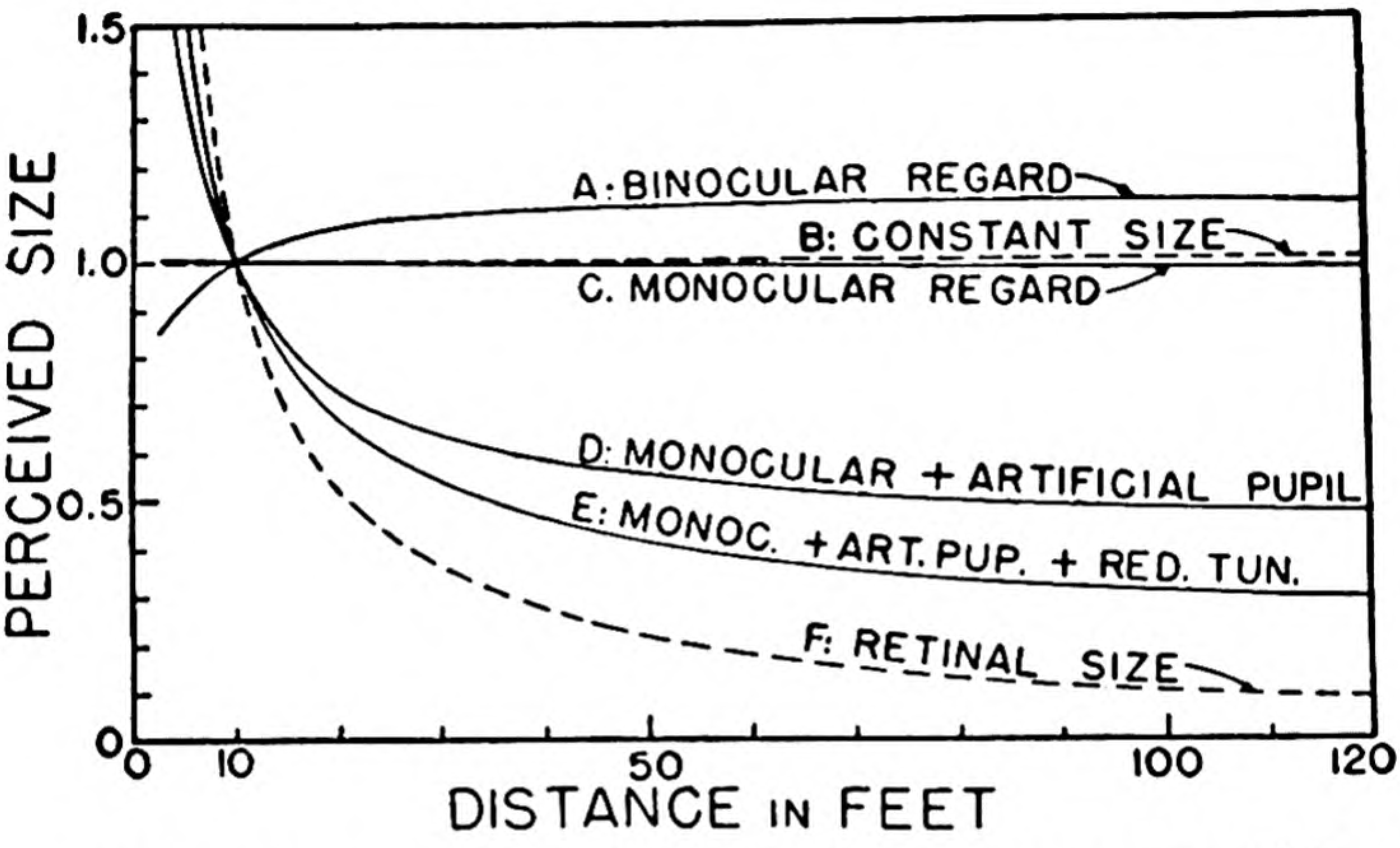


FIG. 64. The Dependence of Size Constancy on Environmental Context. Curves *B* and *F* are theoretical curves representing constant size and retinal size respectively. The other curves represent experimental measurements under the various viewing conditions indicated (From E. G. Boring, The perception of objects, *Amer. J. Physics*, 1946, *14*:104, by permission of the journal and author.)

conditions, however, perceived size decreases as a function of distance, and the empirical function approaches the line defined by the law of the visual angle.

The functions of Fig. 64 clearly demonstrate that it would be erroneous to speak of a general law of size constancy. Variations in perceived size as a function of distance depend upon a set of parameters such as those illustrated in the above experiment. There are, of course, other parameters which are important. Several in-

vestigators have found that perceived size varies with the subject's attitude toward the stimulus objects. A naïve, uncritical attitude favors constancy. Furthermore, the meaning and familiarity of the objects play an obvious part in the perception of size. Finally, there has been a persistent search for correlations with such individual differences as age, sex, intelligence, and occupation. These studies have shown considerable individual variation in susceptibility to size constancy but have not as yet led to any stable generalizations.

The Size of Afterimages. The dependence of perceived size on distance, size of the retinal image, and the context provided by the field conditions is well illustrated by the apparent size of afterimages. If you successively project a clear negative afterimage upon homogeneous surfaces at various distances, you will notice a striking change in the size of the afterimage: the farther away the projection surface, the larger the image will appear. In fact, if the projection surfaces are homogeneous fields, the size of the afterimage is proportional to the distance of the surface from the eye. This generalization is known as *Emmert's law.*

At first, this result may come as a surprise. Should not objects grow smaller as they recede into the distance? A more careful analysis will show, however, that Emmert's law is strictly in accordance with the phenomenon of constancy. The apparent size of the afterimage depends upon the size of the retinal patch and the distance at which the image is viewed. The size of the retinal patch is, of course, constant. Therefore, as distance is increased, a larger and larger object would be required to produce that constant retinal patch. Of course, Emmert's law works immediately and automatically, but, in order for the law to operate, *both* size of the retinal image and the distance of projection must be taken into account. Neither the size of the retinal image alone nor the distance alone provides the requisite information. If the size of the retinal image alone were operative, the afterimage would have to remain the same size at all projection distances. If distance alone were the basis for the variation in size of afterimages, the afterimage should diminish in size with increasing distance. It is the joint effectiveness of the size of the retinal image and the information about distance that leads to Emmert's law.

The importance of the information about distance can be dra-

matically illustrated by "misinforming" the perceiver about relative distances in the field. It is easy to prepare a perspective line drawing of a corridor. If you project an afterimage upon the opening to the tunnel, the afterimage will look smaller than if projected at the far end of the perspective drawing. The linear perspective used in such line drawings is, of course, an effective clue to distance. As such, it becomes one of the determinants of the apparent size of the afterimage. It is, therefore, only in terms of the size of the retinal image and the *perceived* distance that the apparent size of afterimages can be explained.

Experiment XII

Discrimination of Visual Depth

Purpose. To measure the accuracy with which an individual can judge visual depth.

Materials. A vertical black screen with a horizontal slot about 1 inch wide is mounted at the end of a table. The slot serves as an aperture through which the subject views the stimulus objects. At a distance of about 5 feet, a homogeneous white surface is erected. Two vertical rods are mounted side by side so that the visual field of the subject includes them. One of these uprights is stationary somewhere midway between the two screens. The subject can move the other upright back and forth in the line of regard by means of strings and pulleys. A meter stick is laid along the track of the adjustable rod and is used to record the subject's judgments. An eyepatch is necessary to secure monocular vision, in one part of the experiment.

Procedure. The subject is seated some distance away from the aperture so that he sees a horizontal opening with two vertical rods appearing in it. The distance between the stationary upright and the eyes of the subject should be held constant. The subject adjusts the distance of the variable upright until it seems to be in the same frontal-parallel plane as the stationary one. After each judgment, the experimenter records in millimeters the distance between the two planes in which the rods are localized. After a few practice trials, ten adjustments are made from a starting position in which the variable rod is set much farther than the standard and ten adjustments from a starting position much nearer to the subject than the standard. The starting positions are alternated from trial to trial.

The above procedure is followed for each of the following conditions:

1. *Binocular vision without movement of the head.* Under this condition, the subject is instructed to hold his head as still as possible while making his judgments.
2. *Monocular vision without movement of the head.* One eye is covered with the eyepatch and, as before, the subject holds his head as still as possible.
3. *Monocular vision with head movement.* One eye is still covered by a patch, but this time the subject is permitted to move his head. The subject should frequently move his head from side to side in making the adjustment, thus utilizing the information available from monocular parallax.

Treatment of Results. The data are arrayed in a table showing the deviation in millimeters between standard and variable rods. If the variable is adjusted too far away, a deviation is given a positive sign. If too close, a negative sign. In this table, the data should be broken down as is shown here.

	Binocular		Monocular Without Movement		Monocular With Movement	
	Approaching	Receding	Approaching	Receding	Approaching	Receding
1.						
.						
.						
.						
10.						

The means of each of the columns should be computed and also the means for each of the three conditions. Some measure of variability should be calculated. These measures provide the answers to the following questions:

1. Is binocular vision superior to monocular in the discrimination of depth?

2. Do head movements aid in the monocular perception of depth?

3. Is there a systematic difference between the approaching and receding series?

Consideration should also be given to the following points: What clues to depth have been eliminated and what clues have been utilized in the judgments? What variables should be further controlled to improve the experiment? What other psychophysical procedures could be used?

EXPERIMENT XIII
SIZE CONSTANCY

Purpose. To determine under several viewing conditions the variations in perceived size as a function of distance.

Materials. A set of square pieces of black cardboard whose sides vary from 2 centimeters to 10 centimeters in steps of 1 centimeter. These squares serve as *reference stimuli*. The sides of the *test square* are 8 centimeters. Each of these squares is mounted on a large white sheet of cardboard which provides a background for the stimulus. A meter stick may be used to measure the distances of the stimuli from the observer. An eyepatch is to be used to cover one eye in order to obtain monocular viewing. Finally, a tube made of black material is employed as a reduction tunnel.

General Experimental Procedure. The experiment is conducted in a long room or hallway relatively free of furniture or other objects. The subject is seated 1 meter away from the point at which the reference stimuli are presented. The reference squares are presented at the side of the subject so that he has to turn his head about 45 degrees in order to fixate the stimulus. Directly in front of the subject the test square is first presented at a distance of 1 meter. The experimenter then presents the series of reference stimuli both in *ascending* and *descending* orders. The subject is instructed to indicate which of the reference squares appears equal to the test stimulus. Several determinations, say, ten, should be made at each distance.

The test stimulus is then successively moved to a distance of 2, 4, and 8 meters. At each of these distances, the matches are obtained as before. In addition to *increasing* the distance of the test stimulus from the subject, it is desirable to *decrease* it successively, starting with determinations at 8 meters and ending with determinations at 1 meter. Thus, there is a *receding* and an *approaching* series.

Viewing Conditions. The procedure just outlined is used under the following three conditions:

1. *Unimpeded binocular vision.* The subject is allowed to view the stimuli without restrictions.
2. *Monocular vision.* One of the eyes is covered by an opaque patch.
3. *Monocular vision with reduction tunnel.* One eye is covered as in condition 2, and in addition, the subject is required to view the stimuli through a reduction tunnel.

Treatment of Results. The data obtained are the average sizes of the reference stimuli judged to be equal to the test stimulus at each of the

distances under the various conditions of observation. The data may be conveniently grouped as shown in the accompanying table.

	Approaching Series		Receding Series		
Distance	Ascending	Descending	Ascending	Descending	Av.
1					
2					
4					
8					

The entries in the body of this table are the lengths of the sides of the reference squares matched to the test squares.

Overall averages are taken by rows to give a representative value for each of the distances. These averages are then plotted against distance. On the same graph are entered the theoretical functions defined by the law of the visual angle and by perfect size constancy. The question which this experiment has attempted to answer is: Where, between these extremes, do the empirical functions fall as the viewing conditions are altered?

It is possible to isolate the effects of receding, approaching, ascending, and descending series of presentation by plotting separate functions for each of these conditions.

General References

Boring, E. G., *Sensation and perception in the history of experimental psychology,* New York: D. Appleton-Century Co., 1942, Chap. 8.

Carr, H. A., *An introduction to space perception,* New York: Longmans, Green and Co., 1935.

Woodworth, R. S., *Experimental psychology,* New York: Henry Holt and Co., 1938, Chap. 26.

The Spatial Framework

Ewert, P. H., A study of the effect of inverted retinal stimulation upon spatially coördinated behavior, *Genet. Psychol. Monogr.,* 1930, *7*:177-363.

Koffka, K., *Principles of Gestalt psychology,* New York: Harcourt, Brace and Co., 1935, Chaps. 6 and 7.

Stratton, G. M., Vision without inversion of the retinal image, *Psychol. Rev.,* 1897, *4*:341-360; 464-481.

Wooster, M., Certain factors in the development of a new spatial coördination, *Psychol. Monogr.*, 1923, 32, No. 4.

The Perception of Distance

Monocular Determinants

Boring, E. G., *Sensation and perception in the history of experimental psychology*, New York: D. Appleton-Century Co., 1942, pp. 263-271.

Boring, E. G. (ed.), *Psychology for the armed services*, Washington: The Infantry Journal, 1945, pp. 35-39.

Binocular Parallax and Stereoscopy

Boring, E. G., *Sensation and perception in the history of experimental psychology*, New York: D. Appleton-Century Co., 1942, pp. 282-288.

Boring, E. G. (ed.), *Psychology for the armed services*, Washington: The Infantry Journal, 1945, pp. 40-46.

Carr, H. A., *An introduction to space perception*, New York: Longmans, Green and Co., 1935, Chaps. 6, 7, and 8.

Southall, J. P. C., *Introduction to physiological optics*, New York: Oxford University Press, 1937, pp. 236-262.

Woodworth, R. S., *Experimental psychology*, New York: Henry Holt and Co., 1938, pp. 654-665.

Convergence and Accommodation

Baird, J. W., The influence of accommodation and convergence upon the perception of depth, *Amer. J. Psychol.*, 1903, *14*:150-200.

Boring, E. G., *Sensation and perception in the history of experimental psychology*, New York: D. Appleton-Century Co., 1942, pp. 271-282.

Carr, H. A., and Allen, J. B., A study of certain relations of accommodation and convergence to the judgment of the third dimension, *Psychol. Rev.*, 1906, *13*:258-275.

Swenson, H. A., The relative influence of accommodation and convergence in the judgment of distance, *J. Gen. Psychol.*, 1932, 7:360-380.

The Perception of Size

Size Constancy

Boring, E. G., The perception of objects, *Amer. J. Physics*, 1946, *14*:99-107.

Brunswik, E., Distal focussing of perception: size constancy in a representative sample of situations, *Psychol. Monogr.*, 1944, *56*:No. 254.

Holway, A. H., and Boring, E. G., Determinants of apparent visual size with distance variant, *Amer. J. Psychol.*, 1941, *54*:21-37.

Koffka, K., *Principles of Gestalt psychology,* New York: Harcourt, Brace and Co., 1935, 87-96; 235-240.

Sheehan, M. R., A study of individual consistency in phenomenal constancy, *Arch. Psychol.,* 1938, No. 222.

Thouless, R. H., Phenomenal regression to the real object, *Brit. J. Psychol.,* 1931, *21*:339-359; 1931, *22*:1-30.

The Size of Afterimages

Boring, E. G., Size constancy and Emmert's Law, *Amer. J. Psychol.,* 1940, *53*:293-295.

»»» 10 «««

PERCEPTION OF MOVEMENT

OUR discussion of the perception of space and form has been confined to stationary objects. The space in which we move contains, however, at most times a variety of moving objects: people walking on the streets, cars speeding along the highway, birds flying in the air, and so on. The perception of movement, then, is part of the perception of objects in space. Many conditions determining the perception of stationary objects are also operative in the perception of movement. In this chapter we shall examine the types of movement which are perceived and the variables of which they are functions.

Physical Movement and Perceived Movement

We must carefully distinguish the definition of physical movement and of movement as a perceptual phenomenon. The location of a body in physical space is described in terms of its momentary position, its velocity, and its acceleration. The position of an object is determined, of course, by referring it to some standard geometric system of reference. The velocity of an object refers to the direction and rate at which it travels through space. Two physical dimensions are required to specify velocity: distance and time. Acceleration is change in velocity. These are physical concepts, and there are methods for specifying them without direct reference to the perception of motion. Position, velocity, and acceleration can, then, be the independent variables in the experimental study of movement perception. The dependent variables are, of course, the subject's responses to the changes in these physical factors.

There is by no means a simple one-to-one correspondence between the occurrence of physical movement and its perception. Physical movement may occur but remain imperceptible, as in the

case of the hour hand of the clock. Moreover, movement may be perceived under certain conditions when the physical objects are stationary. Movies, it is well known, do not move but are just a rapid succession of stills which we see as moving.

THE PERCEPTION OF A MOVING STIMULUS

Thresholds of Perceived Movement. As in the case of other perceptual phenomena, there is a threshold which must be reached before a moving stimulus is seen as moving. First of all, the physical movement must attain a certain minimum velocity in order to be detected. There is no one minimum velocity which can be specified as *the* threshold for movement. This value will vary with the size of the moving object, its visibility, the intensity of the illumination, and the part and state of adaptation of the retina stimulated. For example, other things being equal, movement can be detected more readily with the fovea than the periphery of the retina. Again, other things being equal, the movement of a well-illuminated object can be seen more quickly than that of a dimly illuminated one. We assume, of course, that such comparisons are made with the stimulus at a constant distance from the eye. Under the best viewing conditions, i.e., with foveal stimulation by a clear, well-illuminated object placed 2 meters from the eyes, a velocity of six-thousandths of a foot per second (0.2 centimeters per second) leads to just detectable movement. This value gives a good idea of how slow a moving object can be seen to move. There is not only a minimum but also a maximum velocity limiting the perception of movement. When the velocity exceeds a certain speed (e.g., 44 feet per second at 2 meters), a blur or flicker rather than a moving object is seen.

The Effect of Distance. The farther away an object is, the more slowly does its image move across the retina. If we assume constant physical velocity, the arc through which the retinal image moves in a unit time becomes smaller and smaller as the movement occurs at a greater and greater distance. We would expect, then, that the object would appear to move more and more slowly, the farther away it is. Here again, however, we find a discrepancy between the amount of physical change and the amount of change perceived. It is true that the farther an object, the more slowly it seems to move. But the decrease in perceived velocity is not pro-

portional to the decrease in the arc through which the retinal image moves per unit time. Parallel to size constancy, there is a certain amount of *movement constancy.*

The Relativity of Perceived Movement. As in the case of size judgments, it would be difficult to obtain an unequivocal answer to the question of how fast an object appears to move. We may get an estimate in terms of miles per hour, but we would not know the size of the units of the subjective scale of speed. It is necessary to resort to comparisons between a reference stimulus and a test stimulus. Such comparisons will provide us with information about the parameters on which changes in perceived movement depend. The following experiment, a classic in this field, provides us with a paradigm for the investigation of the conditions of the perception of physical movement.[1]

The velocity of a test stimulus was adjusted by the subject until it seemed equal to that of the reference stimulus, whose physical velocity was held constant. More specifically, a band of paper with figures on it moved past an aperture. In the case of the reference stimulus, this band moved with constant velocity. In the case of the test stimulus, the velocity of the band was adjusted by the subject. Both apertures were located at the same distance. Various features of the test stimulus were systematically varied, and the resulting changes in the velocity matches examined. As an example of the results obtained, let us describe one of the findings.

To study the effects of relative linear size on perceived velocity, all of the linear dimensions of the test stimulus were halved, as were the size of the moving figures and the distance between successive figures. Such changes had a striking effect upon the perceived velocity. The reference stimulus moved with a velocity of 10 centimeters per second. When the test stimulus moved 5.3 centimeters per second, it seemed to move as fast as the reference stimulus. The dimensions of the test stimulus were halved, and its physical velocity needed to be reduced by a factor of two in order to preserve the subjective equality of the two velocities! It is interesting to note that when only some of the dimensions of the test stimulus were halved, the reduction in physical velocity was less

[1] J. F. Brown, The visual perception of velocity, *Psychol. Forsch.*, 1931, *14*: 199-232.

than one-half. The structure of the visual field in which the movement occurs clearly is an important condition of perceived velocity. Features other than size were varied. It was found, for example, that when the brightness of the test stimulus was increased, its physical velocity had to be increased in order to maintain a constant perceived velocity. The reader can undoubtedly think of many other variables whose influence on perceived velocity might be tested. The influence of such variables could easily be investigated by means of the same experimental procedure.

Afterimages of Movement. Just as continued exposure to a stationary object results in a visual afterimage, so continued perception of a moving stimulus produces an afterimage of movement. If we gaze at a waterfall for a period of time, say, a minute or so, and then turn our regard to the bank on the opposite side, the side of the bank will appear to move upward, i.e., in the direction opposite to that of the waterfall. This afterimage of movement may be termed negative because of this reversal of the direction of movement. If we look out of the window of a moving train at the ties of an adjacent track, we see them moving in the direction opposite to the train. When the train comes to a sudden halt, we now see the ties moving forward, i.e., the direction of the afterimage is reversed. Furthermore, we may feel as if we ourselves were moving backward. This feeling of backward motion is not just due to the negative afterimage of movement. Our static sense also exhibits phenomena which are parallel to the negative visual afterimage. Aftereffects that show a phenomenal reversal seem to be common to several types of sensory action.

Apparent Movement in the Absence of Physical Movement

How is physical movement translated into stimulation that gives rise to perceived movement? The answer to this question is by no means known, but it is clear that a moving object successively stimulates discrete, adjacent receptors in the retina. The distribution of these receptors is punctiform; they are separated from each other by tiny spaces, and a continuously moving stimulation in the strict sense of the term is impossible. There must be some inte-

grating process which makes possible the perception of continuous movement.

The perception of movement on the basis of discrete stimulation of different parts of the retina is even more pointedly illustrated by the phenomena of so-called apparent movement. Of course, all perceived movement is apparent, but this phrase has been reserved for those special cases in which the stimulus objects remain stationary and movement is perceived.

Optimal Movement and Phi. When two discrete stimuli, e.g., two bars of light, are presented in rapid succession, one of several types of apparent movement may be perceived. Which kind of movement is seen depends upon a variety of factors among which the following are most important: the distance between the two stimuli, their intensity, and the time interval between the first and second stimulus. In order to illustrate the various types of movement, let us hold constant all factors except the time interval between the first and second stimulus.

Our subject is seated before a screen in a dark room. In this screen are two slots, each about an inch wide and 4 inches apart. These slots can be illuminated in rapid succession by means of separate light sources behind them. We flash the light first through the left slot (*L*) and 15 milliseconds later, we flash the other light through the right slot (*R*). The subject reports that he sees simultaneously two bars of light. For time intervals up to about 20 milliseconds, the report of simultaneity is given. As the interval is further increased, the subject begins to see apparent movement. With a time interval of about 60 milliseconds, there is what is called *optimal movement*. The left bar appears to jump over to the right one. It is as though the bar itself moved. (Naturally, if *R* is presented before *L*, the movement is from *R* to *L*.) Let us further increase the time interval between the two flashes. If we make the interval as long as 200 milliseconds, bars appear in succession and without movement. There is no direct transition from optimal movement to succession. At some time interval between 60 and 200 milliseconds, *pure phi movement* is often seen. The movement seems to be divorced from any object, and the subject may describe it as "pure movement" or perhaps as a "gray flash."

Korte's Laws of Movement. For purposes of illustration, we

have held constant all factors except the time interval. Variations in the distance between the bars or in the intensity of illumination affect the readiness with which apparent movement is seen. The relationships among the three variables, time, intensity, and distance, have been worked out and are known as *Korte's laws.* Although these laws hold only over a limited range of values, they still serve to specify in a quantitative manner the conditions of apparent movement. The interrelationships among the three variables may be stated as follows:

1. If the intensity is held constant, the time interval for optimal movement varies directly with the distance between the stimuli.
2. If the time interval is held constant, the distance for optimal movement varies directly with the intensity.
3. If the distance between the stimuli is held constant, the intensity for optimal movement varies inversely with the time interval.

Let us illustrate statement 1. Suppose that optimal movement has been obtained for a given set of values of the three variables. If the time interval between the flashes is now slightly increased, the movement may be destroyed. It can, however, be brought back by increasing the distance of separation between the bars. The other statements may be interpreted in a similar manner.

Stroboscopic Movement. The kinds of movement which we have just described—the perception of movement without a moving stimulus—have also been termed *stroboscopic movement.* Originally, this term was applied to the movement that results when slightly different pictures are presented in quick succession. The principle upon which motion pictures depend is the best example of stroboscopic movement. As we know, the apparent movement is brought about by the projection at a fairly rapid rate (about 24 frames per second) of a series of still pictures. Each of these stills presents a slightly different view of the individual or object which is then seen as moving on the screen. Animated cartoons utilize the same principle.

Other Kinds of Apparent Movement. In our previous discussion, we have assumed that the intensities of the two stimuli were the same. An interesting phenomenon of movement develops if the second flash is made considerably brighter than the first. Under these conditions, the direction of the movement may be reversed.

If a dim flash on the left is followed by a bright flash on the right, the movement may be from right to left. Sometimes this reversed movement is followed immediately by a movement back to the position of the second and brighter flash. This phenomenon of reversed movement due to a brightness difference is known as *delta movement.*

One more type of apparent movement will be mentioned. Only one stimulus object is required for its demonstration. If the illumination of a figure in the visual field is suddenly increased, the figure may be seen to expand at first and then to contract to a stable size. Similarly if this bright illumination is suddenly decreased, the figure may be seen to contract. This expansion or contraction of an object due to changes in illumination is designated as *gamma movement.*

Several other varieties of apparent movement have been reported in the psychological literature, but we shall not stop here to enumerate them. Instead, we wish to reiterate that the kind of apparent movement depends upon the particular conditions of stimulation. We have already indicated the importance of intensity, time interval, and distance between successive stimuli. There are other important factors. The size of the discrete stimuli and the length of time during which each is exposed are examples of additional variables. Equally important as these stimulus variables is the attitude that the observer adopts during the experiment. A naïve, uncritical attitude favors the perception of apparent movement. An analytical and critical approach may well detract from the impressiveness of the phenomenon. It is interesting to note the parallel between the effect of attitude upon the degree of constancy and of apparent movement. A naïve, uncritical attitude tends in both instances to enhance the phenomenon.

Auditory and Tactile Stimuli to Perceived Movement

Auditory as well as tactile impressions may contribute to the perception of movement. When a tone is led separately to each ear but so that there is a phase difference between the two stimuli, a phenomenon termed *auditory movement* occurs. A phase difference, it will be recalled, is an important condition for the localization of sounds. The tone is localized in different positions depending upon

the degree of the phase difference. The perceived tone can be made to move by varying continuously the phase difference between the stimuli. If the sound is at first localized on one side of the head, it appears to move around to the other side. Sometimes the subject reports that the sound travels through his head. As the phase difference between the tones reaches 180 degrees, the sound may jump back to the side from which it started. The experience comes as a surprise to the listener and has been described as the hearing of a *phantom sound.* Auditory phenomena paralleling visual phi have also been reported. The quick successive presentation of discrete tones or clicks may result in apparent auditory movement.

When a stimulus is continuously moving along the skin, the subject experiences it as moving. Let us recall that the distribution of cutaneous receptors is punctiform and that the occurrence of tactile movement must result from an integrative process. Again, apparent movement may be obtained without actually moving a stimulus along the skin, but by stimulating in rapid succession two discrete points. The investigators reporting this phenomenon agree, however, that tactile phi is less compulsory and occurs less frequently than visual phi.

There is a question as to whether or not there is pure auditory or tactile movement. Visual imagery and interpretation in terms of visual space undoubtedly play an important part in the experience of these phenomena.

Autokinetic and Induced Movement

The perception of movement is strongly influenced by the spatial framework in which the stimulus is seen. When a spatial framework is lacking, a stationary point of light may appear to move. If you carefully fixate a pinpoint of light in an otherwise completely dark room, you will, after a while, see the point of light drift to one side or the other *(autokinetic movement).* There are wide individual differences in the direction and degree of movement, but the phenomenon of movement occurs for virtually all observers. In the dark room, there is no stationary visual coördinate system to which the single point of light may be referred. For most subjects, the light makes rather small excursions, but on occasion it may deviate from a central position as much as 40 degrees. As we shall see in Chapter

20, this movement is subject to considerable modification, and the phenomenon has been particularly suitable for the investigation of social influences on this perception.

The influence the spatial framework has on the perception of movement is further illustrated by the phenomena of *induced movement*. When there are only two light objects in a homogeneous dark field, and one of these is in physical motion, movement will be seen, but it does not necessarily follow that it is the objectively moving object that will be seen to move. What is seen depends not only upon the relative displacement of the two objects, but also upon some of the factors of spatial organization. Suppose one of the objects is the outline of a geometrical design enclosing another. Even if it is the larger figure that is objectively moving, the smaller one will be seen to move. The larger figure serves as a framework for the smaller thing that moves. The framework, being more stable, is seen at rest.

Experiment XIV
Apparent Movement

Purpose. To study the dependence of optimal movement upon intensity and upon the time interval and distance between two discrete stimuli.

Materials. There are many ways of arranging apparatus for the study of apparent movement. One convenient setup is the following. Two small boxes, each with a single slot about one inch wide and 6 inches long cut in the front surface, contain the sources of illumination (light bulbs). The slots are covered with milk glass. A rheostat is used for altering the intensity of the lights. The time interval between the onset of the two lights is automatically controlled by an interval timer. Such a timer can employ a disk driven at a constant speed with contact switches located on its circumference. The contact switches should be such that the lights are on for a very brief interval. There are also, of course, electronic interval timers. The distance is controlled by the position of the light boxes.

Procedure. The experiment should be conducted in a dark or dimly lighted room. The subject is seated about 6 feet from the two light boxes which are placed on a table. The distance between the boxes and the intensity of the lights are set to some arbitrary values (say, with the slots 6 inches apart and the lights fairly bright). The time interval between the two lights is, then, systematically varied from about 20 milliseconds to about 200 milliseconds. After each presentation of the two lights, the subject reports what he saw. When he sees the one bar of light moving

over to the other, he sees optimal movement. The time intervals at which simultaneity, optimal movement, pure phi, and successivity occur, are noted. The intensity of the lights is then reduced and this procedure repeated. Finally, the distance between the two boxes is varied and the time interval of optimal movement is again determined. If time permits, it is desirable further to test Korte's laws, stated on p. 211, by holding the time interval constant at, say, 60 milliseconds, and simultaneously varying the distance and intensity to obtain optimal movement.

Treatment of Results. For each combination of intensity and of distance between the bars, state the time intervals at which each of the types of apparent movement is seen. Compare the results obtained with the predictions made upon the basis of Korte's laws. Note especially whether or not pure phi (movement without an object) is more readily seen under one condition than another.

General References

Boring, E. G., *Sensation and perception in the history of experimental psychology,* New York: D. Appleton-Century Co., 1942, Chap. 15.

Boring, E. G., Langfeld, S. H., and Weld, H. P., *Psychology: a factual textbook,* New York: John Wiley and Sons, 1935, Chap. 11.

Perception of a Moving Stimulus

Thresholds and Relativity of Perceived Movement

Brown, J. F., The visual perception of velocity, *Psychol. Forsch.,* 1931, *14*:199-232.

Brown, J. F., and Voth, A. C., The path of seen movement as a function of the vector field, *Amer. J. Psychol.,* 1937, *49*:543-563.

DeSilva, H. R., An analysis of the visual perception of movement, *Brit. J. Psychol.,* 1929, *19*:268-305.

Koffka, K., *Principles of Gestalt psychology,* New York: Harcourt, Brace and Co., 1935, pp. 282-304.

Afterimages of Movement

Bowditch, H. P., and Hall, G. S., Optical illusions of motion, *J. Physiol.,* 1882, *3*:297-307.

Apparent Movement

Optimal Movement and Phi

DeSilva, H. R., An experimental investigation of the determinants of apparent visual movement, *Amer. J. Psychol.,* 1926, *37*:469-501.

Dimmick, F. L., An experimental study of visual movement and the phi phenomenon, *Amer. J. Psychol.*, 1920, *31*:317-332.

Higginson, G. D., The visual apprehension of movement under successive retinal excitations, *Amer. J. Psychol.*, 1926, *37*:63-115.

Wertheimer, M., Experimentelle Studien über das Sehen von Bewegung, *Z. Psychol.*, 1912, *61*:161-265.

Korte's Laws

Korte, A., Kinematoskopische Untersuchungen, *Z. Psychol.*, 1915, *72*:193-206.

Other Kinds of Apparent Movement

Boring, E. G., *Sensation and perception in the history of experimental psychology*, New York: D. Appleton-Century Co., 1942, pp. 596-597

Auditory and Tactile Stimuli to Movement

Burtt, H. E., Auditory illusions of movement—a preliminary study, *J. Exper. Psychol.*, 1917, *2*:63-75.

Burtt, H. E., Tactual illusions of movement, *J. Exper. Psychol.*, 1917, *2*:371-385.

Stevens, S. S., and Davis, H., *Hearing: its psychology and physiology*, New York: John Wiley and Sons, 1938, Chap. 6.

Whitchurch, A. K., The illusory perception of movement on the skin, *Amer. J. Psychol.*, 1921, *32*:472-489.

»»» 11 «««

EXPERIMENTAL ANALYSIS OF JUDGMENT

RESPONDING to objects in our environment often involves making judgments about them. Many situations contain elements of doubt and uncertainty. When we settle a doubt or resolve an uncertainty, we make a judgment. Sometimes our judgments express the type of fine discrimination which we study in psychophysical experiments. We decide whether or not a weak stimulus is present, such as a faint tone or light; we judge two stimuli as same or different. More often than not, the judgments which we must make do not call for minute discriminations of this kind nor do we always judge such clearly defined attributes as loudness or brightness. On innumerable occasions, we judge situations as strange or familiar, threatening or reassuring, interesting or dull, and in terms of many other qualities. We must decide whether an action is ethical or unethical, useful or futile. Daily we must judge people as good or bad, able or incompetent, friendly or hostile. Such judgments continually determine our course of action and are verified or disproved by the results of our actions.

The Tasks of Analysis

The experimental analysis of judgment faces three major tasks: (1) an adequate description of the stimuli judged and of the situation in which the judgment is made; (2) an analysis of the characteristics of judgments, such as their variability, the speed and confidence with which they are made; and (3) the establishment of lawful relationships between the properties of the stimulus situation and the characteristics of judgments. In searching for such relationships, the experimenter should not forget that it is a person with well-established habits, values, and needs who makes the

judgments. No account of judgment can ever be complete which fails to take account of such factors "in the person."

TYPES OF JUDGMENT

Virtually anything may be the object of a judgment: a physical object, an abstract thought or proposition, any event, past, present, or future, actual or hypothetical. A classification of judgments according to the object judged would, therefore, be difficult, cumbrous, and at best, of little practical use. It is more profitable to attempt a systematic classification according to the *type of response* which mediates the judgments.[1]

Perceptual Judgments. Judgments may reflect the way a subject perceives an object or some characteristic of an object. A tone may be judged loud or soft, high-pitched or low-pitched. A design may be judged symmetrical or asymmetrical, large or small. In all such cases, the subject's judgment is based on a particular aspect or dimension of his perceptual experience. A loud tone may be high or low in pitch. But, in judging loudness, we try to disregard pitch; in judging pitch, to disregard loudness. A large design may be symmetrical or asymmetrical. In judging symmetry, we disregard size; in judging size, we disregard symmetry. In making any one simple perceptual judgment, we concentrate on a given characteristic of the stimulus object to the exclusion of its other characteristics which are not relevant to the judgment.

Affective Judgments. An affective judgment expresses an individual's like or dislike for an object, its pleasantness or unpleasantness for him, in short, its personal value. In judging a tone loud or soft, we should pay no attention to the fact that hearing the tone may be a pleasant or an unpleasant experience. In judging a design as symmetrical or asymmetrical, the fact that symmetry may be "pleasing to the eye" is irrelevant. In affective judgments, it is precisely this kind of personal reaction which is made explicit. A large variety of stimulus materials have been used in experiments on affective judgment: foods, drinks, and odors as well as words and pictorial materials. Typically, it has been the subject's task to rate

[1] For the classification of judgments presented here, we are indebted to Donald M. Johnson, A systematic treatment of judgment, *Psychol. Bull.*, 1945, *42*: 193-224.

such stimuli on a scale of pleasantness-unpleasantness, or acceptance-rejection, ranging from maximum acceptance, through indifference, to maximum rejection. Subjects usually find it quite easy to make such judgments and are fairly consistent with themselves from one occasion to another.

Affective judgments are also often obtained in the measurement of social and political attitudes. In attitude questionnaires, we are often asked to rate various national or social groups in the order of our preference. We are asked to indicate whether certain types of people are acceptable to us as neighbors, colleagues, or friends. In essence, such judgments are affective judgments since they are based on our personal values and inclinations rather than on an attempt to gauge the "objective" situation as accurately as possible. It is important to point to the continuity of the processes underlying the type of affective judgment studied in the laboratory and the type of judgment which springs from our political and social attitudes.

Affective judgments are readily made, but often the judge would be at a loss to state what specific characteristics of the stimulus object are responsible for his acceptance or rejection, for experienced pleasantness and unpleasantness. The results of past conditioning and learning and deep-seated motives all come into play, often in subtle ways of which the subject is unaware. Through systematic variation of the stimulus materials and intensive study of his subjects, the experimenter may sometimes gain at least partial insight into the determinants of affective judgments.

Conceptual Judgments. Frequently, judgments are based on abstract or conceptual characteristics of the stimulus objects, characteristics which are derived from some scheme of classification. A botanist classifying plants or a zoologist classifying animals assigns specimens to various categories by judging such characteristics as manner of reproduction or skeletal structure. A scientist evaluating theories according to their comprehensiveness or consistency, or a logician judging propositions according to their conformance with formal rules, provides another example of conceptual judgment. Finally, many of our judgments of personality are of the abstract or conceptual type when we attempt to evaluate people in terms of such traits as coöperativeness, conscientiousness, or liberalism. To

make such judgments, we usually need some "yardstick," some ready-made categories in terms of which we judge the concrete cases which we encounter. For example, we may use a scale of honesty, ranging from scrupulous uncompromising honesty to deliberate deceit, and try to place any given individual somewhere along this scale or "continuum." Conceptual judgments are difficult to make because of their very abstractness. Often, what we believed to be a conceptual judgment may turn out, after searching analysis, to be an affective one.

Conceptual Judgment and Thinking. Conceptual judgments are processes which are frequently considered under the heading of "thought," for successful thinking often depends on the ability to abstract relevant features from a complex situation and on the use of classificatory schemes. The analysis of thought, in its broad sense, cannot, of course, be equated with the analysis of conceptual judgments. The interdependence of thought processes with perceptual functions, learning, memory, and transfer of training must be considered. In this book we shall limit ourselves, however, to analyses of these basic cognitive functions which come to their full fruition in what we call thought. Our understanding of thought will depend on the thoroughness of our grasp of the basic cognitive processes.

The Expression of Judgments

There are two general ways (by no means mutually exclusive or incompatible with each other) in which we can study another person's judgments: (1) verbal report, and (2) inference from nonverbal behavior.

Verbal Report. We can ask our subject to report his judgment verbally, expressing it either orally or in writing. Our subject will readily inform us whether he judges one tone louder than another, whether an odor is pleasant or unpleasant, whether he considers a person honest and coöperative. With proper instructions and training, most subjects are able to refine their judgments and make quantitative distinctions among different amounts of the characteristic judged. They can assign different degrees of pleasantness and unpleasantness to taste or odor, indicate varying degrees of agreement or disagreement with a political statement, use a gradu-

ated scale of honesty or coöperativeness in their evaluation of people, and so on. Such verbal reports have been used very widely in the experimental investigation of judgment.

Inference from Nonverbal Behavior. A verbal or written report is, however, only one way in which a judgment can be expressed. Without asking any questions or receiving any reports, we can observe an individual's behavior and infer from his behavior what judgments he has made. Suppose we confront a hungry individual with two foods. He eagerly eats one of the foods but does not touch the other, even withdraws from it or pushes it away. His behavior expresses a judgment of preference: he likes one food better than another. We infer his judgment from the choice he makes. In this sense, we can study the "judgments" of organisms which are incapable of verbal communication. As long as subjects make consistent choices among alternative stimulus objects, they tell us their judgments as clearly as they could by any verbal statement.

There is no inherent contradiction between these two methods of studying judgment. It would be a serious error to say that in studying verbal reports we deal with "mere words," whereas the other method is concerned with "real behavior." Verbal reports are as much *behavior* as are muscular movements. In man, so much problem-solving behavior—and judgment is a type of problem solving—is mediated by words and symbols that verbal report provides at least one important avenue of approach to the study of judgment. Critics have sometimes raised the question of how sure we can be that a person expressing a judgment is "telling the truth." He may hold one judgment and express another, perhaps in order to please the experimenter. But the same criticism may be applied to inference based on nonverbal behavior. An individual may pursue one course of action although he may prefer another. Our conclusion is that we can never do more than study overt behavior. Such behavior may be verbal or nonverbal, and it is dangerous to consider one type of behavior as inherently more important than another. Each has its place in the experimental analysis of judgment.

Stimulus Scales and Response Scales

We have stated as the general aim of the experimental analysis of judgment the establishment of lawful relationships between the

properties of the stimulus situation and the characteristics of judgments. Wherever possible, we should like to state these relationships in quantitative terms. To achieve this goal, we need (1) quantitative indices for the description of the stimulus objects, and (2) quantitative indices of the variations in judgments. In short, quantification depends on the availability of proper *stimulus scales* and *response scales*. Let us illustrate the concepts of stimulus scale and response scale by means of a simple example.

The stimulus objects are five lines of different lengths, say, 2, 4, 6, 8, and 10 inches. These lines are presented to the subject in random order, and he is instructed to judge each line as *long* or *short*. At first the subject will be guessing, since he does not know what range of lengths the series comprises, but after one or two rounds of stimulus presentations, his judgments will settle down. He will almost invariably call the 2-inch line *short* and the 10-inch line *long*. His judgments of the 6-inch line will probably be variable, and he will call it *long* part of the time and *short* part of the time unless he is making a constant error of overestimation or underestimation. The series of lines ranging from 2 to 10 inches is the stimulus scale. The responses, *long* and *short*, constitute a simple, two-category response scale. Furthermore, it is possible to state a quantitative relationship between the two scales. We can estimate that length of line which is likely to yield 50 percent of judgments of *long* and 50 percent of judgments of *short*,[2] i.e., the threshold or "boundary" separating the categories *long* and *short*.

We may now require our subject to make finer distinctions in his judgments and to use three categories of response: *long*, *short*, and *medium*. The 2-inch line would still be called *short* almost invariably, and the 10-inch line, *long*. The 6-inch line would probably be called *medium* on the majority of the trials. The 4-inch line would probably be called *medium* part of the time and *short* part of the time; similarly, the judgments of the 8-inch line would be divided between *medium* and *long*. While the stimulus scale remained the same, the subject's response scale has been refined and now includes three categories. The threshold of each of these categories can be computed.

[2] The computational procedure for estimating a category threshold is described in Chapter 2 in connection with the method of constant stimuli.

The judge's task may be rendered even more exacting. We may ask him to use the integers from 1 to 5 in judging the lines, calling the shortest 1, and the longest 5. After some training, a subject can perform this task successfully and assign the numbers to the different lines with a high degree of consistency. Again the thresholds of the various categories can be calculated. The stimulus scale is still the same, but the response scale has been refined even further and now comprises as many response categories as there are stimulus items. A response scale results whenever numbers (or other symbols) are assigned to stimulus objects in accordance with perceived changes in magnitude. Clearly, response scales are highly flexible. To a given stimulus scale, there may correspond a variety of response scales, ranging from two-category judgments to a refined scale requiring many and difficult discriminations on the part of the judge.

The Relativity of Judgments

The experimental procedure which we have illustrated in the preceding paragraphs is known as the *method of single stimuli.* The method derives its name from the fact that each stimulus to be judged is presented singly, without a standard stimulus. The method of single stimuli is representative of many judgments which we have to make in daily life. Frequently we do not judge objects in pairs but rather judge them one at a time. We see a movie and we call it interesting or dull without comparing it with some standard movie. We call a person's voice loud or soft without comparing it to a standard voice level, and so on. But even though we may be judging one object at a time, such judgments are, nevertheless, relative to a response scale which we have acquired in the course of daily living. We have seen a large number of movies and so have acquired a personal response scale for rating them, ranging from very interesting to very dull. We have heard a great many voices and, again, have acquired a personal response scale, ranging from very soft to very loud. When we judge any single object, we place it, often without making an explicit comparison, somewhere along the range of similar objects which we have experienced. We then assign to it a value along the personal response scale which we have for this type of stimulus object. In making our judgments, we measure things with a subjective yardstick (personal response scale), a yardstick

which we acquire as a result of being exposed to a range of stimulus objects.

Such subjective yardsticks are highly flexible, and response scales readily shift, expand, and contract with a change in the context in which the judgments are made. Suppose we change our series of lines so that the series now consists of lines 10, 12, 14, 16, and 18 inches long. We again require our subject to use two categories of response—*long* and *short*. As soon as he is acquainted with the series, he will call the 10-inch line *short* on most trials. But in the previous series, he usually called this very same line *long*. Obviously, the context in which this line is judged has been radically altered; it occupies an entirely different position in the stimulus series. *At any moment in time, the effective response scale is determined by the nature and range of stimuli to which the judge is responding*. Recall again the type of judgments which we continually make about objects around us. A book is small, and a man is large. But if a house is large, then a man is small. And if a book is small, and a house is large, then a man is of medium size.

All judgments, then, are, in the last analysis, comparative judgments. Sometimes the comparisons are made explicit, as in psychophysical experiments which provide the subject with a standard stimulus on every trial. Often the comparisons are not explicitly made as in the method of single stimuli, in which case each stimulus object is evaluated in terms of its position in the series to which it belongs.

The Anchoring of Judgments

Let us return for a moment to our experiment on the judgment of lines. The stimuli are again 2, 4, 6, 8, and 10 inches in length. A subject has learned to assign the numbers from 1 to 5 to them, with satisfactory consistency, calling the shortest line 1, and the longest, 5. We now introduce a 15-inch line into the series and instruct our subject to consider this line as representing the category 5. We then resume the experiment and ask our subject to assign the numbers 1 to 5 to each line presented singly as before. Of course, he will now virtually always call the 15-inch line 5, but his judgments of the other lines will be affected as well. In general, he will be inclined to assign *lower* numbers to the different lines than he did before.

The 8-inch line, for example, which previously had been judged 4 most of the time, will now frequently be called 3. Similarly, the 6-inch line will be called 2 more often than previously. In relation to the 15-inch line, all the other lines are likely to appear shorter than they did before. If, instead of a 15-inch line, we had added a line ½ inch long to the series, the judgments would have been pulled in the opposite direction: to most of the lines *higher* numbers would have been assigned than before this addition to the series. A 4-inch line would then be likely to be called 3 where it had usually been called 2, an 8-inch line would frequently be assigned the number 5, and so on.

Thus, the response scale can be altered appreciably by a change in the stimulus scale. Expressing it a little differently, we can say that the response scale is *anchored* to the stimulus scale, and that a change in the stimulus scale leads to a reanchoring of the response scale. The reponse categories available to the subject (say, the numbers 1 through 5) must be distributed among the various stimuli in the series. When the stimulus series is changed, the responses must necessarily be redistributed. The redistribution of responses has to be such as to encompass the new stimulus as well as the old ones. The response categories may have to cover a wider range of stimuli. In our experiment, the numbers 1 to 5, which had been used for a range from 2 to 10 inches, had to be redistributed to cover a range from 2 to 15 inches or from ½ inch to 10 inches. As a result, the relative position of each of the old stimuli was changed, making them appear shorter in one case, longer in the other. A stimulus which is added to a series and leads to a redistribution of judgments is known as an *anchoring stimulus*. The redistribution of judgments resulting from the introduction of an anchoring stimulus is called *anchoring*.

Anchoring is a general principle of judgment and has been demonstrated with a variety of stimulus materials: judgments of weights, of the inclination of lines, of the length of temporal intervals as well as aesthetic and affective judgments have all been shown to be subject to anchoring. Invariably, it has been possible to change distributions of judgments by adding stimuli to a series and, thus, redefining the meaning of the response categories. Indeed, it is not always necessary actually to add anchoring stimuli to a series. Some-

times it is sufficient to ask a subject to hold an anchoring stimulus in mind while making his judgments. Instead of showing a 15-inch line to our subjects, we could have asked them to imagine a 15-inch line as defining the category 5, and a redistribution of judgments would have occurred similar to that resulting from the actual presentation of the anchoring stimulus.

Anchoring experiments provide a dramatic illustration of the relativity and flexibility of judgments. A response scale is always anchored to the stimulus scale to which it is applied. Changes in the stimulus scale cause the response scale to be reanchored and, consequently, the response categories to be redistributed among the stimulus objects. Truly, there is nothing absolute or fixed about any judgment.

Some General Principles of Judgment

In spite of the great variety of stimulus situations which have been used in the experimental analysis of judgment, certain general principles have emerged. We have already emphasized the relativity of all judgments as well as the universality of anchoring phenomena. There are other principles which appear to be characteristic of the process of judgment in general, regardless of the particular stimulus objects and the particular attributes which are judged. Some of these general principles refer to properties of response scales, others focus on the way in which the judge goes about making his judgments, the nature of his performance. Here are some of these general principles.

The Central Tendency of Judgment. In many situations subjects show a pronounced tendency to avoid extremes in their judgments, to use the center of their response scale more heavily than the extremes. In judging sizes of a series of objects, for example, the extremely small ones are likely to be overestimated, and the extremely large ones to be underestimated. Similarly, when the lengths of time intervals are judged, the extremely short ones are judged too long, the estimates of the extremely long ones are too short. The judgments are displaced from the extremes toward the center. This fact is aptly described as the *central tendency of judgment*. The central tendency of judgment has been demonstrated with a variety of stimulus materials, including judgments of personality variables.

In judging intelligence, we hesitate to call an individual a moron or a genius and settle for a safer and less colorful rating. When moral judgments are involved, judges may feel an even greater need to shrink away from extremes.

The Round-Number Tendency. Response scales differ in the number of categories which they include. Some response scales, as we have seen, comprise only two categories (e.g., *long* and *short*); others include much larger numbers of categories, e.g., all the numbers from 1 to 10, 1 to 20, and so on. When judgments have to be expressed by assigning numbers to stimulus objects, subjects show clear preferences for some numbers over others. They will habitually favor round numbers, such as 5, 10, 15, 20, etc., or even numbers. This tendency—the *round-number tendency*—can be observed in many practical situations. In grading their students on a scale from 0 to 100 teachers will show it, as will judges deciding on the exact length of a criminal's sentence when the law prescribes only a certain range. Nor are scientists immune from it when reading instruments and graphic records. The round numbers provide, as it were, clear-cut landmarks along the response scale. Preference for them is continually reinforced by daily usage. Such habits of response readily assert themselves whenever and wherever judgments have to be made.

Variability at Different Points of the Response Scale. The same judgment is not always given in response to the same stimulus (see Chapter 2). The greater the scatter of responses to a given stimulus, the greater the variability of the judgment. When a series of stimuli is judged (such as in our experiment with lines of different lengths), judgments of stimuli near the center of the series typically show greater variability than judgments of stimuli near the extremes of the series. Even though the *physical* distances between successive stimuli are equal (e.g., lines of 2, 4, 6 inches, etc.), it seems that *subjectively* the distances near the extremes are larger than they are toward the center of the scale. Hence, discriminations near the center are more difficult and more variable.

The Halo Effect. When faced with the necessity of making a difficult judgment, we may substitute a general impression for a precise discrimination. We may invoke, wittingly or unwittingly, irrelevant information to help us make the judgment. Judgments of

specific personality characteristics are often determined in this manner. When asked to evaluate such characteristics of a person as honesty, friendliness, reliability, and so on, each of our judgments may be influenced by a general impression which we have formed of that person; or our knowledge of *one* characteristic may color our judgments of all others. We may know from experience that a person is honest and, thereby, be predisposed to rate him high not only in honesty, but also in friendliness, reliability, and other characteristics as well. When a generalized attitude determines the judgments of specific characteristics, we speak of the *halo effect*. The halo effect has been observed most frequently in judgments of personality. The halo effect is especially pronounced when characteristics carrying moral approval or disapproval are judged. In general, whenever a subject is required to judge vague or poorly defined characteristics, his judgments may become subject to the halo effect. One should not conclude too hastily, however, that there is a halo effect just because a judge assigns similar ratings on a variety of characteristics. Only if it can be shown on independent evidence that such uniformly high (or uniformly low) judgments are invalid, can a halo effect be reasonably inferred.

The Atmosphere Effect. The atmosphere effect is closely related to the halo effect and exemplifies another way in which a judgment may be determined by a general impression rather than by an exact discrimination. In judging the validity of syllogisms, for example, subjects may be guided not only by formal logical considerations but by the "atmosphere" created by the premises. When one of the premises is negative, e.g., "no *x*'s are *y*'s or "some *x*'s are not *y*'s," a negative atmosphere is created, and subjects are inclined to accept a negative conclusion. Similarly, the presence of a "particular" premise, i.e., a statement such as "some *x*'s are *y*'s" predisposes subjects to favor a conclusion of the same type even though this conclusion be logically invalid. As in the case of the halo effect, the judgment depends on a general attitude toward the stimulus objects. The close similarity between the halo effect in judging personality and the atmosphere effect in logical reasoning serves to emphasize the basic continuity of the processes which operate in judgment and in reasoning or "thinking."

Judgment Time, Confidence, and Difficulty of Judgment.

One characteristic of judgment which varies systematically with changes in the stimulus situation is the *speed* with which the judgment is made. Of course, every judgment takes a certain amount of time. But *how much* time elapses between the presentation of a stimulus and the subject's report is not accidental or haphazard. Indeed, speed of judgment may be more sensitive to changes in the stimulus situation than the subject's report. The experimental work on the relationship between judgment time and other characteristics of judgment has given rise to a number of generalizations:

1. The greater the difficulty of the discrimination, the longer is the judgment time. This fact can easily be demonstrated in a conventional psychophysical experiment. If the perceived difference between the standard stimulus and the comparison stimulus is large, the judgment (*larger* or *smaller*) is usually given quite rapidly. If, on the other hand, the difference between the two stimuli is small or zero, the judgment is usually given much more slowly. The same general relationship between difficulty and speed of judgment can be shown in other judgment situations as well. Suppose, for example, a worker has to classify the products of a factory as either meeting specifications or as defective. Those items which are clearly defective or clearly superior will be classified quickly. It is the borderline cases which will be judged most slowly. In general, those stimulus objects which fall near the threshold of a response category will be judged more slowly than those which are clearly above or below the threshold.
2. Correct judgments are made more quickly than incorrect judgments. This relationship follows necessarily from the fact that judgments of stimuli near category thresholds require the longest times. Judgments of stimuli in the threshold area are the most variable and are more likely to be wrong than judgments of stimuli which appreciably exceed the threshold value. These variable, and frequently wrong, judgments require long times as compared with less variable, and usually correct, judgments.
3. The more confident a subject is about his judgment, the shorter the judgment time. When a judgment is difficult, e.g., when a very fine discrimination is required, the subject not only makes his judgment slowly, but he is also likely to have little confidence in his judgment, certainly less confidence than in a judgment

which is easily and quickly made. Thus, there is an inverse relationship between judgment time and the degree of confidence which the subject expresses. This relationship has been frequently established in experimental investigations.

The Reliability and Validity of Judgments

Reliability. Reliability refers not to the correctness or incorrectness of a judgment but to the consistency with which a judgment is given in response to a specific stimulus object. If a given stimulus object always evokes the same judgment, we have complete reliability. At the other extreme, if the judgments vary as widely as they would if the judges were guessing at random, we have complete lack of reliability. In most practical situations, neither of these extreme conditions is found. Few judgments have 100 percent reliability; nor do we usually encounter zero reliability. In most cases, judges show a fair amount of consistency which safely takes their performance out of the realm of random guessing.

In the analysis of judgments, we may be concerned with two types of reliability: (1) the reliability of the judgments made by one judge and (2) interjudge reliability.

In the case of a single judge, we can inquire how consistent he is in his responses to a set of stimuli. To answer this question, we can require him to repeat his judgments on two (or more) occasions and correlate the two (or more) sets of judgments. If the correlation is high, his judgments have high reliability. This method of testing reliability sometimes runs into difficulties, for the judge may *remember* his previous ratings (say, in judging a group of individuals on some personality trait) and the test becomes in part a test of memory rather than a check of consistency.

The problem of reliability of judgments can be approached from a somewhat different point of view. Taking some attribute or characteristic (say, again, some personality trait), we can question how consistent different judges are with each other in their judgments. Reliability may then be determined by correlating the judgments of different individuals. If the correlation is perfect, we have complete interjudge reliability. If the correlation is zero, the reliability is nil. If there is satisfactory agreement between the judges, we can use their *mean* judgment as a better estimate of the attribute in ques-

tion than either of the judgments taken singly. Sometimes, when highly abstract or complex judgments are involved, it is advisable to average the responses of many more than two judges.

Validity. A judgment is valid to the extent that it is independently verified by the use of an *external criterion.* In many instances validation must come from the observation of behavior. For example, if a personnel officer judges a prospective employee to be coöperative and conscientious, his judgments can be validated against the worker's subsequent conduct. To the extent that predictions made on the basis of his judgment come true, his judgment is proved valid. Since each judge is fallible, at least to some extent, the pooled judgment of several individuals will usually be more valid than the judgment of a single individual. Thus, not only the reliability but also the validity of judgments increases with the number of judges.

Often it is difficult to find an external criterion for the validation of judgments. Aesthetic and affective judgments especially are difficult to validate. As the old adage has it, one man's meat is another man's poison, and likes, dislikes, and preferences cannot be right or wrong. It is possible, however, to consider a judge's actions as criteria of validity. To the extent, then, that a judge's actions, e.g., the choices he makes, are in accord with the affective and aesthetic judgments he expresses, these judgments may be presumed to be valid.

It is important to note that highly reliable judgments may be invalid. A judge or group of judges may be perfectly consistent in a judgment which is proven invalid by subsequent events. Validation depends entirely on the extent to which an outside criterion bears out the assertion made in the judgment.

Experiment XV

The Anchoring of a Response Scale

Purpose. To demonstrate (1) the acquisition of a scale of responses to a series of stimuli, and (2) the changes in the response scale caused by the introduction of an anchoring stimulus.

Stimulus Materials. A variety of stimulus materials can be used, provided they can be arranged to form a series increasing in magnitude in discrete steps. For example, a series of lines of different lengths can be used, a series of weights, a graded series of squares, etc. Since an experiment on the judgment of lines has already been discussed above, we shall

now illustrate the procedure using a series of weights. A convenient series would consist of the following weights: 60, 80, 100, 120, and 140 grams. All the weights should, of course, be indistinguishable from each other by external appearance. A rotating table is a useful device for the presentation of the weights, but they may also be presented by hand.

Experimental Procedure. The experimental procedure falls into three parts: (1) an initial practice series; (2) the training series, during which the subject establishes a scale of responses to the series of weights; and (3) several anchoring series, during which the response scale is modified by the introduction of anchoring stimuli falling outside the original series of weights. The weights are presented and judged one by one.

1. *Practice series.* For the practice series, the subject is instructed as follows:

   "I shall present to you five different weights in random order. I want you to judge the heaviness of each weight, using the numbers 1 to 5. Call the lightest weight 1, and the heaviest weight 5, and assign the other numbers in their proper order to the intermediate weights. As soon as you have lifted the weight, tell me the number which you wish to assign to it. At the beginning you will, of course, be guessing, but you will soon learn to discriminate the different weights."

   With a series of five weights, a practice series of twenty-five trials should be sufficient. Thus, each weight is presented five times in random order.

2. *Training series.* After the initial practice series, a training series is given. This series may consist of 100 trials, with each weight presented twenty times in random order. The subject is instructed to continue making his judgments as before.

3. *Anchoring series.* Following the training series, the anchoring stimuli are introduced. It is useful to run at least two anchored series, one with a large anchoring stimulus (heavier than any weight used in the series) and one with a small anchoring stimulus (lighter than any weight used in the series). If time permits, it is, of course, possible to use more than two anchored series. The order in which different anchors are introduced should be balanced. If there are two anchoring stimuli, one large and one small, half the subjects should receive the large anchor first; the other half should begin with the small anchor. If there are more than two anchored series, similar schedules of rotation can be worked out.

   Suppose a weight of 180 grams is chosen as the heavy anchoring stimulus. On each trial, the subject *first* lifts the anchor weight and then

one of the weights from the regular series which he continues to judge as before. The subject is given the following instructions:

> "In the series which follows, you are to continue judging the weights as before. From now on, however, you will be given *two* weights to lift on each trial. The first weight will be a standard which will remain constant on all trials. Use this first weight to define the category 5, i.e., the standard will represent a weight to which the number 5 is to be assigned. The second weight will be one of those you have judged before. Continue to assign numbers to them just as you did in the preceding series. Remember, use the first weight merely as a standard or reference and always judge the second weight."

Choosing a weight of, say, 20 grams, as the lighter anchor, the instructions given to the subject are identical with those given for the heavy anchor series. In this case, however, the anchor defines the category 1, rather than the category 5, and 1 is substituted for 5 throughout the instructions.

Treatment of Results. The experimenter records each judgment on a prepared score sheet. For each of the series, he then prepares a *distribution of judgments* which may look somewhat as is shown in the accompanying table.

Stimulus Weight	Frequency of Judgment				
	1	2	3	4	5
60					
80					
100					
120					
140					

Thus the frequency with which the numbers 1 to 5 were assigned to each stimulus weight is recorded. It is now possible to determine the *threshold* for each of the categories. For a given category, the threshold is, of course, defined as that stimulus value which will be assigned to that category 50 percent of the time. Suppose a weight of 80 grams was called 3 on 20 percent of the trials, a weight of 100 grams was called 3 on 70 percent of the trials. Hence, the threshold of the category 3 lies between 80 grams and 100 grams. By linear interpolation, we locate this value at 92 grams. In a similar manner, thresholds for all the categories are determined for all the series.

The important result to look for is the *shift* in category threshold caused by the introduction of anchoring stimuli. It is to be expected that introduction of a heavy anchor will tend to raise the category thresholds,

especially in the upper parts of the scale, i.e., of those categories nearer to the anchor. Conversely, the introduction of a light anchor should result in a general lowering of the category thresholds, especially in the lower part of the scale, i.e., again in that part of the scale which is closest to the anchor. *The extent of the shift is a quantitative index of anchoring.* To measure this shift, we compare the category thresholds obtained under the three conditions: training, large-anchor series, and small-anchor series. If several anchors are used at each end of the scale, the magnitudes of the shifts caused by them should also be compared.

There is an alternative way for treating the data. Taking the stimulus weights in turn, we determine for each of them the mean category value assigned to it in a series. Each of these means would be based on twenty cases (or whatever the frequency with which each stimulus was present in a series). Anchoring can then be gauged by the upward or downward change in the mean category assigned to each of the stimuli. The reliability of such shifts can be determined by conventional tests of statistical significance.

The above description illustrates the procedure of an anchoring experiment, using weights as stimulus materials. No basic changes in the procedure would result from the use of other stimulus materials.

Experiment XVI

Judgment Time as a Measure of Difficulty of Decision

Purpose. This experiment is concerned with the demonstration of the general principle that judgment time increases as the threshold of a response category is approached.

Materials. Again we may choose from a large variety of stimulus materials. Any series of stimuli is adequate which leads to decisions of *varying degrees of difficulty*. For illustrative purposes, we choose as stimulus materials two sets of geometrical figures:[3] a learning series and a test series.

The learning series consists of five angles, the sizes of which are 60, 70, 80, 90 and 100 degrees. The test series consists of fifteen angles, covering the range from 10 to 150 degrees in successive 10-degree steps. Thus, the test series comprises the original figures plus ten new angles, five of which are smaller and five larger than those used in the learning series. Each of the angles is drawn in heavy black ink on white cardboard.

Experimental Procedure. The learning series is presented to the subjects with the following instructions:

[3] This experiment is modeled after D. Cartwright, Relation of decision time to the categories of response, *Amer. J. Psychol.*, 1941, 54:174-196.

"I shall present to you a series of angles of varying size. I want you to look at them carefully since you will be required later to pick out the ones you have seen from a large number of angles."

The cards are then presented to the subject five times in random order. Each card is exposed for a constant period of time, say, ½ second.

After a short rest, the test series is presented with these instructions.

"I shall again present to you a series of angles. Some of these were included in the previous series, others are new and they have not been shown to you before. Every time a card is shown, I want you to indicate whether or not you recognize it as having been shown in the first series."

The test series is then presented ten times, each time in a new random order. The subject's judgments are recorded as well as the time elapsing between the presentation of the stimulus and the subject's judgment.

Exposure and Timing Devices. If no other facilities are available, the cards may be presented by hand and the judgments timed by means

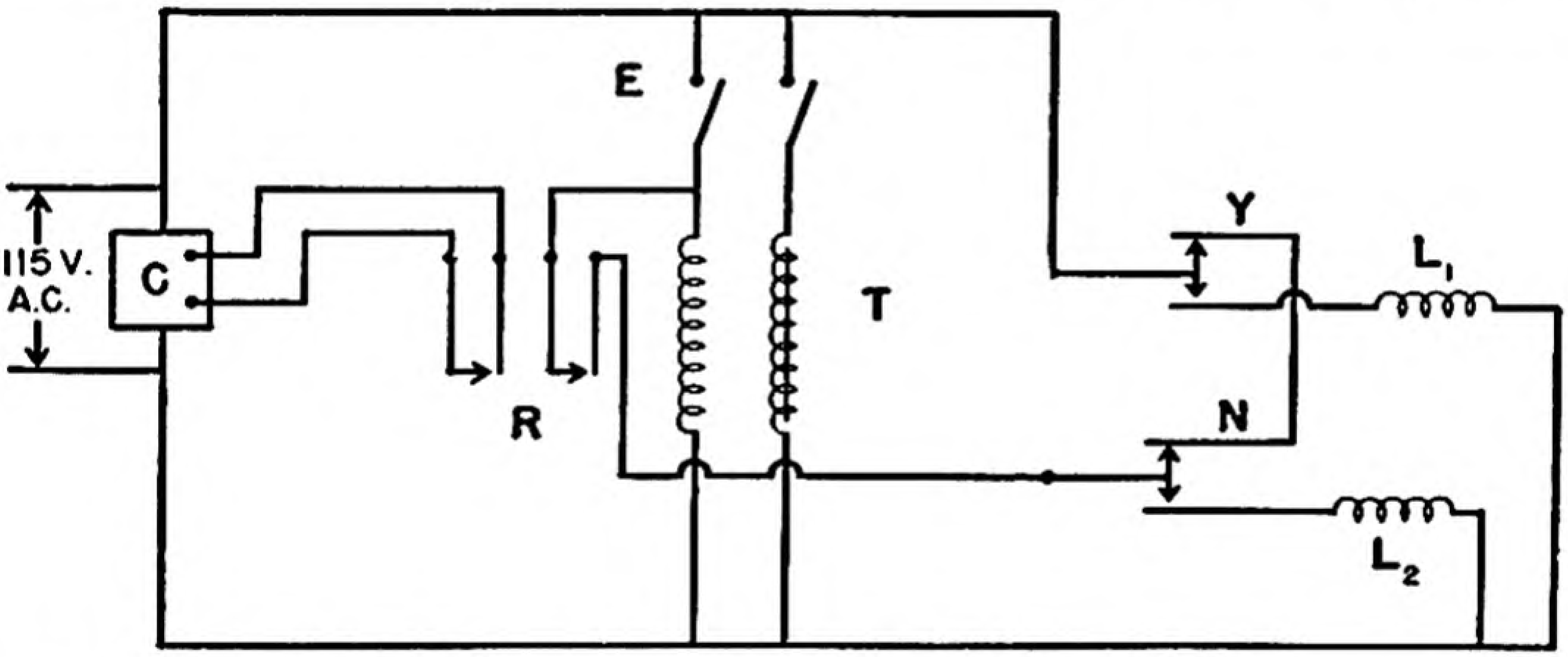


FIG. 65. Wiring Diagram of Apparatus Used in Experiment XVI. The experimenter (E) momentarily closes his double pole, single throw switch, and thereby activates the clock (C) through a double pole, single throw 115-volt relay (R). (This relay is normally open.) At the same time he releases the tachistoscope slide held in place by the magnet at T. The subject presses one of two keys (Y to indicate "Yes" and N to indicate "No"), and thereby stops the clock. By pressing his key, the subject also causes a bulb to light up (L_1 or L_2), enabling the experimenter to record the nature of the response.

of a stop watch. However, the accuracy of the experiment is greatly enhanced if standardized exposure and timing devices are used.

The stimulus cards may be conveniently exposed by means of a

tachistoscope.[4] For the test series, the following arrangement will insure precise timing of the judgments: the tachistoscope is incorporated in an electric circuit wired in such a way that the exposure of the stimulus card (opening of the shutter of the tachistoscope) closes a circuit and activates a chronoscope. Also in the circuit are two keys which the subject uses for indicating his judgments. One of the keys means "Yes," the other, "No." Depression of a key breaks the circuit and stops the clock. The time elapsing can then be read directly from the clock. Each of the response keys is also connected with a small bulb which lights up whenever the key is depressed. Thus, the experimenter can record both the nature and time of the judgment. A schematic wiring diagram showing such a circuit appears in Fig. 65.

Treatment of Results. For each of the test stimuli, we compute the percentage of trials on which it was recognized (more exactly on which it was judged as having been exposed in the learning series) and the average judgment time. Thus, we obtain a distribution arranged as shown in the accompanying table.

Test Stimulus (Angle)	% Recognition	Av. Judgment Time
10°		
20°		
30°		
.		
.		
.		

We then plot the two distributions on the same graph: (1) percent recognitions as a function of stimulus size, and (2) average judgment time as a function of stimulus size. The critical results to watch for are the changes in these dependent variables as the learning stimuli are approached on the left and on the right. At what point is the judgment time maximal? Is the distribution of judgment times unimodal or bimodal, i.e., has it one or two peaks? How are the changes in judgment time related to the changes in frequency of recognition?

In terms of these and similar questions, the experimenter can apply the general principles of judgment discussed above.

[4] A tachistoscope is an instrument used to expose visual stimuli for very brief time intervals. In its simplest form, it consists of a screen with a window, mounted in an upright frame. The stimulus is held in the frame behind the screen and is hidden from view while the screen is stationary. When the screen is dropped, the stimulus is momentarily exposed through the window in the screen as the latter moves past the stimulus card.

General References

Guilford, J. P., *Psychometric methods,* New York: McGraw-Hill Book Co., 1936.

Johnson, D. M., A systematic treatment of judgment, *Psychol. Bull.,* 1945, *42*:193-224.

Stimulus Scales and Response Scales

Hunt, W. A., and Volkmann, J., The anchoring of an affective scale, *Amer. J. Psychol.,* 1937, *49*:88-92.

McGarvey, H. R., Anchoring effects in the absolute judgment of verbal materials, *Arch. Psychol.,* 1943, No. 281.

Rogers, S., The anchoring of absolute judgments, *Arch. Psychol.,* 1941, No. 261.

Volkmann, J., The method of single stimuli, *Amer. J. Psychol.,* 1932, *44*:808-809.

Volkmann, J., Anchoring of absolute scales, *Psychol. Bull.,* 1936, *33*:742-743.

Volkmann, J., Hunt, W. A., and McGourty, M., Variability of judgment as a function of stimulus-density, *Amer. J. Psychol.,* 1940, *53*:277-284.

Wever, E. G., and Zener, K. E., The method of absolute judgment in psychophysics, *Psychol. Rev.,* 1928, *35*:466-493.

General Principles of Judgment

Central Tendency

Hollingworth, H. L., The inaccuracy of movement, *Arch. Psychol.,* 1909, No. 13.

Hollingworth, H. L., Experimental studies in judgment, *Arch. Psychol.,* 1913, No. 29.

Round Number Tendency

Coover, J. E., *Experiments in psychical research at Leland Stanford University,* Stanford University, 1917.

Halo Effect

Allport, G. W., *Personality: a psychological interpretation,* New York: Henry Holt and Co., 1937, pp. 446-448.

Rugg, H., Is the rating of human character practicable? *J. Educ. Psychol.,* 1921, *12*:425-438, 485-501; 1922, *13*:30-42, 81-93.

Thorndike, E. L., A constant error in psychological ratings, *J. Appl. Psychol.,* 1920, *4*:25-29.

Atmosphere Effect

Sells, S. B., The atmosphere effect, *Arch. Psychol.*, 1936, No. 200.

Judgment Time, Confidence, and Difficulty of Judgment

Cartwright, D., Relation of decision-time to the categories of response, *Amer. J. Psychol.*, 1941, *54*:174-196.

Johnson, D. M., Confidence and speed in the two-category judgment, *Arch. Psychol.*, 1939, No. 241.

Seward, G., Recognition time as a measure of confidence, *Arch. Psychol.*, 1928, No. 99.

Volkmann, J., The relation of the time of judgment to the certainty of judgment, *Psychol. Bull.*, 1934, *31*:672-673.

RELIABILITY AND VALIDITY

Guilford, J. P., *Psychometric methods*, New York: McGraw-Hill Book Co., 1936, Chap. 13.

Symonds, P. M., *Diagnosing personality and conduct*, New York: D. Appleton-Century Co., 1931, Chap. 3.

Thurstone, L. L., *The reliability and validity of tests*, Ann Arbor: Edwards Bros., 1931.

»»»» 12 ««««

REACTION TIME AND ASSOCIATION

I. REACTION TIME

ONE of the classical experiments of the psychological laboratory concerns the *speed* with which a response is given to a stimulus, or, as it is technically known, *reaction time*. It was once believed that the study of reaction time would result in a kind of "mental chronometry," that with the aid of this technique it would be possible to measure the duration of such mental processes as choice and discrimination. Some, indeed, hoped to time the functioning of the will. These hopes have remained pious wishes of psychological history, for it has proved impossible to isolate from each other, and to time, the processes of discriminating, choosing, and willing in a complex act. The mental chronometry which the early experimentalists had envisaged never came to pass. Nevertheless, the reaction-time experiment has had a long and distinguished history and is still a standard procedure in the psychological laboratory. Although it has failed to give us a mental chronometry, the reaction-time experiment has proved useful in many other ways. Reaction time has remained a sensitive measure of the *readiness* of an organism to respond to changes in the environment. The degree of such readiness or preparedness varies lawfully with both the nature of the stimulus situation and the state of the organism, its sets and attitudes. To know what determines readiness to respond often is of critical importance in the prediction and control of behavior. The measurement of individual differences in reaction time, moreover, is often useful in the selection of men for jobs requiring accurate timing and sureness of response.

49

Reaction Time, Judgment Time, and Latency

The study of temporal characteristics of response is of major concern in many areas of psychological experimentation. A variety of measures of response speed have been developed. We shall begin, therefore, with a brief survey of such measures and of terminology.

Simple Reaction Time. Reaction time is the period between the onset of the stimulus and the beginning of the response. In the *simple reaction-time* experiment, there is usually a single stimulus such as a light, a sound, or a touch to the skin, and the subject is instructed to react as soon as he can to the stimulus, usually by withdrawing his finger from a telegraph key. In this situation, a premium is put on the sheer speed of response, and the subject is often so highly motivated to react rapidly that he may at times "jump the gun" and give his response prematurely, i.e., before the stimulus has been delivered.

Disjunctive Reaction Time. In the *disjunctive reaction-time* experiment, the subject's task is more complicated. Instead of one stimulus, two or more different stimuli are presented in a random order, e.g., lights of different colors, sounds of different quality, alternations of sound and light, and so on. The subject is instructed to react to one but not to the other stimuli. For example, he may be told to respond to the onset of a green light, but not to that of a red one. Thus, a discrimination is required in order to make the correct response. The task may be rendered even more complex by using not only alternative stimuli but also alternative responses. The subject may be instructed to respond to one type of stimulus (say, green light) with his left hand and to another type of stimulus (say, red light) with his right hand. Under such conditions, two discriminations are involved: identification of the stimulus and identification of the appropriate response or, as the early experimenters liked to put it, a discrimination (between stimuli) and a choice (between responses). Such discriminations and choices can, to be sure, become highly automatized and after sufficient training are not usually verbalized by the subject.

Judgment Time. The task of the subject may, of course, not be limited to the recognition of the presence of a specific stimulus (the appearance of a light, sound, etc.). He may be called on to

make a judgment about the stimulus, about its intensity, its quality, its pleasantness or unpleasantness, its familiarity, to name but a few types of judgment which may be required. In such situations, the time of response may again be determined—the period between the presentation of the stimulus and the moment at which the subject makes his judgment. Here the responses vary from trial to trial; the subject must decide on a response each time a stimulus is presented. The speed of such responses is usually designated as *decision time* or *judgment time*. Judgment time is, of course, an example of reaction time, but the reaction is at a different level of complexity than simple recognition of a stimulus. Judgment times, therefore, cover a much wider range than reaction times to sensory stimuli. Difficult judgments may require many minutes, whereas even the slowest sensory reactions rarely exceed a second. Judgment time is a sensitive measure of the difficulty or complexity of the judgment which the subject is required to make. For a fuller discussion of this measure, see Chapter 11.

Latency. Speed of response is often used to gauge the effectiveness of stimuli to which an organism is exposed. In such cases, the speed with which a response appears is usually designated by the term *latency*. In the field of learning, latency is sometimes taken as an indication of the *strength* of a learned response. The more thoroughly an association between two items is established, for example, the more rapidly one item may be given upon presentation of the other. In motor and verbal learning, a strong association may be distinguished by speed of response. For an illustration of latency as a measure of strength of learning, turn to Fig. 66.

Speed of response may also be an index of sensitivity. The speed with which neural tissue reacts to changes in the environment varies with the sensitivity of the tissue to the particular type of stimulation involved. The response of the optic nerve to various kinds of light stimulation is a case in point. Fig. 67 shows a result obtained in a classical investigation of visual processes in which latency of response was used as an index of sensitivity. The subject here was Mya, a photosensitive crab, whose sensitivity to intensity could not very well be tested by a conventional psychophysical method. As Fig. 67 shows, the latency of Mya's typical response to light (with-

drawal of its antennae) varies inversely with the intensity of the stimulus.

These differences in terminology do not represent systematic dif-

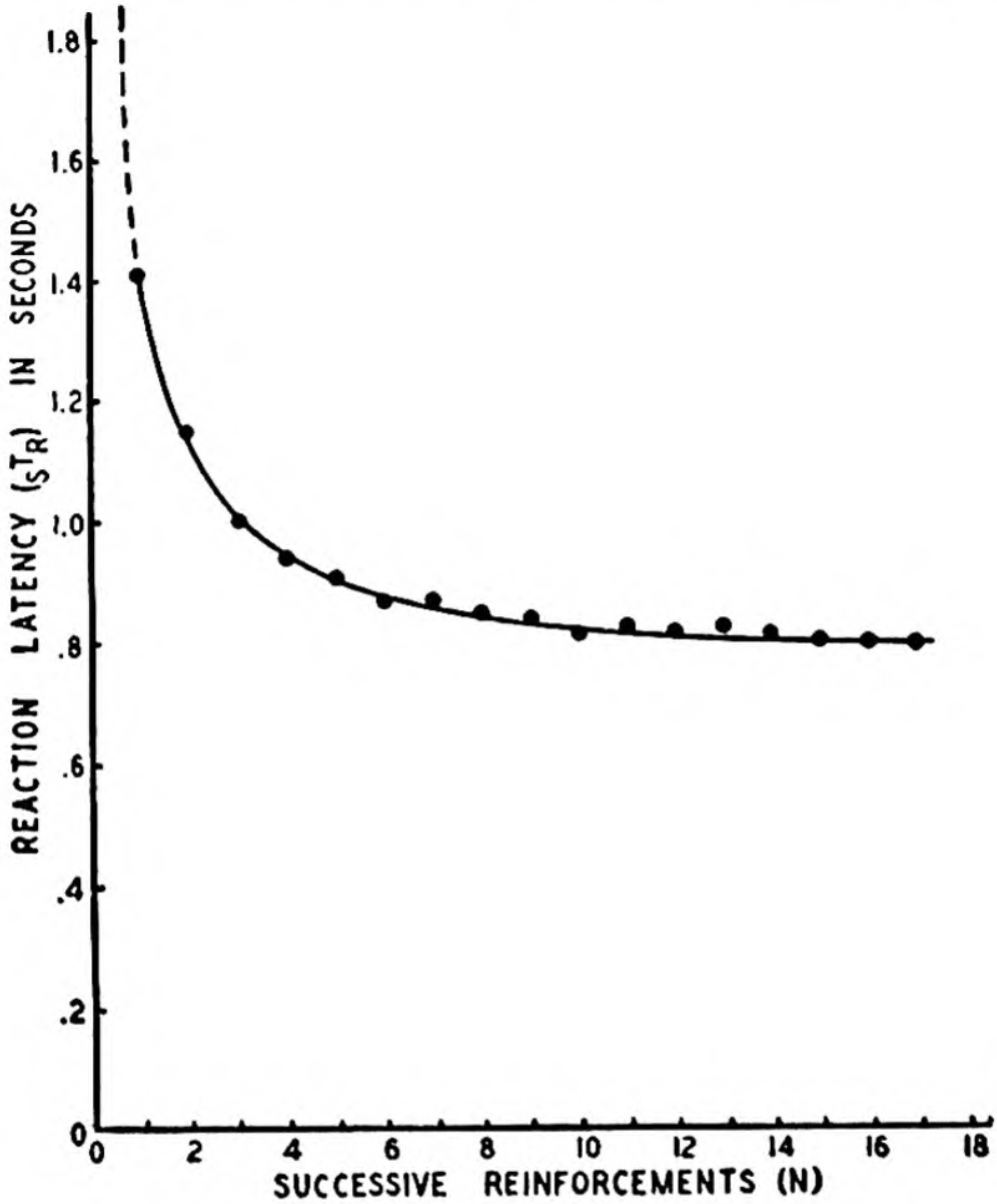


FIG. 66. Latency as a Measure of Response Strength in Learning. This curve shows how the latency of responses in a paired associate experiment decreases as a function of the number of reinforcements (correct responses). (From C. L. Hull, *Principles of behavior,* 1943, p. 105, by permission of Appleton-Century-Crofts, Inc.)

ferences among the various measures of response speed; they have arisen in the interest of convenient reference to different contexts of measurement. Basically, all these methods are continuous: they

are sensitive indices of the organism's readiness to respond to any environmental change.

We now turn to the reaction-time experiment and first consider the standard apparatus used in this work.

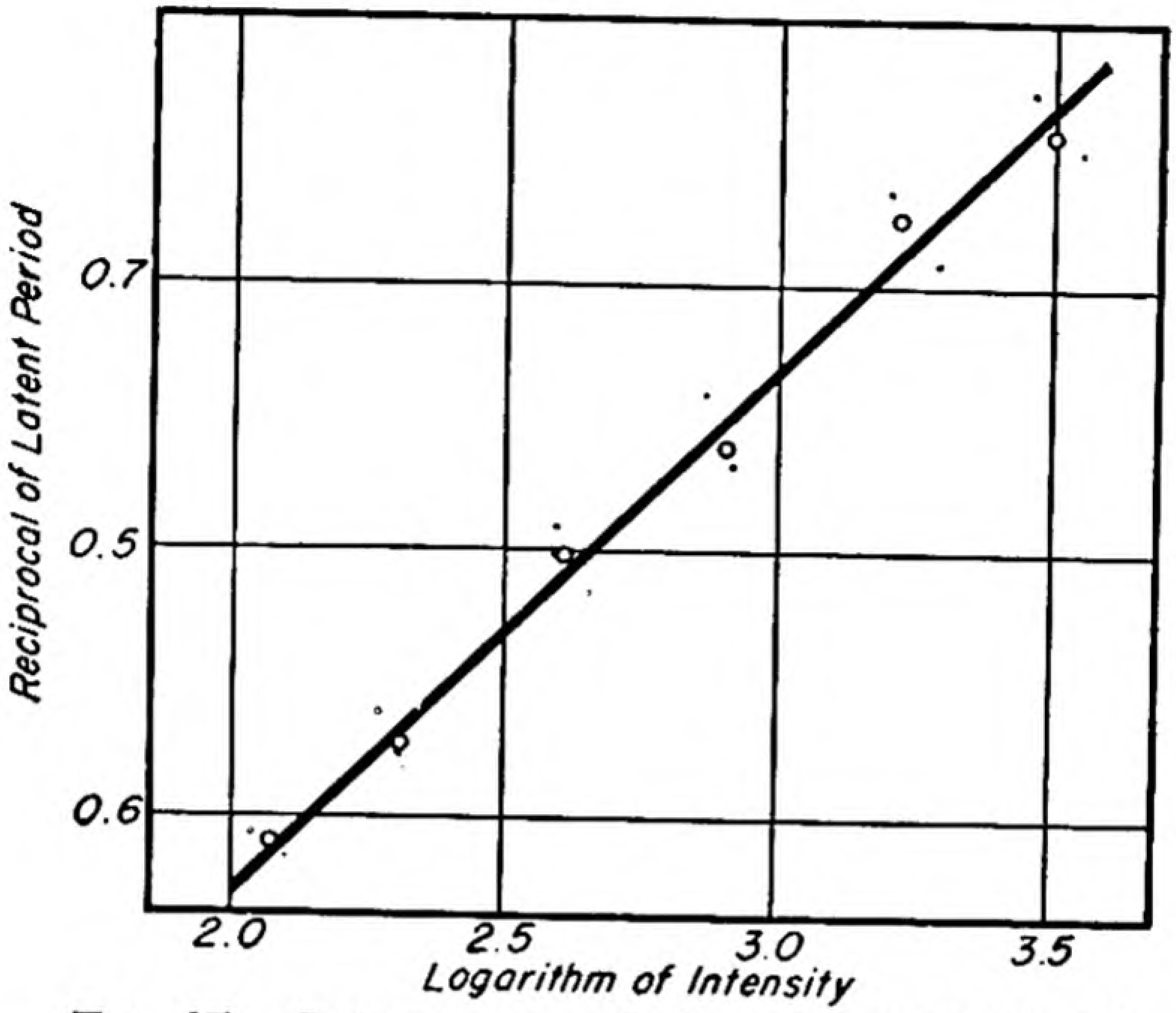


FIG. 67. Latency as an Index of Sensitivity. The greater the intensity of the light, the faster does the organism (the photosensitive crab *Mya*) respond to stimulation. (From S. Hecht, Vision II. The nature of the photoreceptor process. In C. Murchison (ed.), *A handbook of general experimental psychology,* 1934, p. 716, by permission of Clark University Press.)

Standard Apparatus in Reaction-Time Experiments

The reaction-time experiment requires standardized conditions for the delivery of the stimulus, for the subject's response, and for the measurement of the time interval between the onset of the stimulus and the beginning of the response. A typical arrangement consists of a stimulus source (such as a light bulb or buzzer), a pair of telegraph keys and a chronoscope.

The chronoscope is a precision instrument for the measurement of short time intervals. A constant-speed motor moves a pointer

around a dial. (Usually the motor is an electric one, though highly accurate clockwork motors were used in the early work.) The dial is calibrated in hundredths of a second, or even finer divisions. The pointer of the chronoscope is attached to a magnetic clutch, so that it is moved around the dial at a constant speed when the clutch engages with the motor. When the clutch is disengaged, the movement of the pointer is stopped.

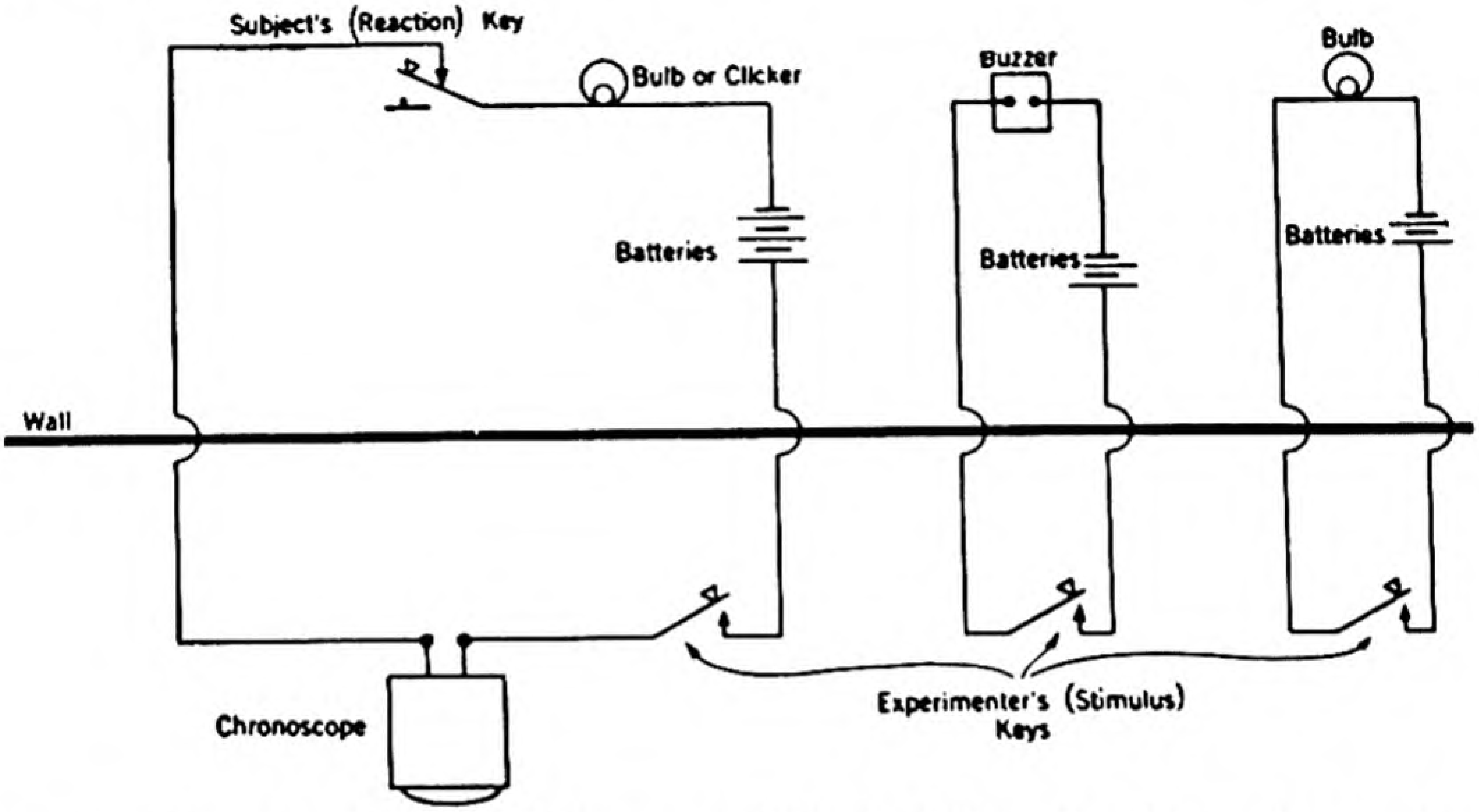


FIG. 68. A Schematic Diagram of Typical Reaction-Time Apparatus. (From N. L. Munn, *A laboratory manual in general experimental psychology*, 1948, p. 1, by permission of Houghton Mifflin Company.)

The two telegraph keys are used to engage and disengage the motor and, thus, to start and stop the movement of the pointer. One of the keys is operated by the experimenter. When he presses his key, he causes the stimulus to be delivered to the subject (e.g., a light to flash or a buzzer to sound) and, at the same time, activates the magnet which engages the motor and thereby starts the movement of the pointer around the dial. The subject operates the second key. Shortly before the onset of the stimulus, he is given a "ready" signal and places his finger on the key. As soon as he can react to the stimulus, he withdraws his finger from the key, breaking the electric circuit and thereby disengaging the motor so that the movement of the pointer around the dial is stopped almost instantaneously.

When stimuli or responses are verbal (such as in the association experiment described below), voice keys instead of telegraph keys are used. The electric circuit is made and broken by vibrations of a diaphragm set in motion by the voice of the experimenter or subject. Thus, the experimenter by speaking into his voice key completes the circuit, activates the magnet and starts the pointer of the chronoscope. When a subject speaks into his key, the vibrations of the diaphragm break the circuit and stop the chronoscope.

A schematic diagram of a reaction-time setup is shown in Fig. 68. As the figure indicates, a variety of sensory stimuli may be used in a reaction-time experiment. Indeed, reaction times have been obtained for stimuli in all the sense modalities, and the modality stimulated is one of the factors with which reaction time varies over a considerable range.

The Determinants of Reaction Time

We now turn to a discussion of the determinants of reaction time, the variables of which its magnitude and variation are a function. For purposes of analysis, we shall distinguish three sets of variables: (1) the characteristics of the stimulus, (2) set and attitude of the reactor, and (3) individual differences among subjects.

Reaction Time as a Function of Stimulus Characteristics

Differences Among Sense Modalities. In the conventional reaction-time experiment, the subject is instructed to respond to the stimulus by withdrawing his finger from the key as soon as possible. The sense modality stimulated is clearly an important parameter of every reaction-time experiment. As the accompanying table of typical reaction times indicates, the speed of response differs considerably for the various sense modalities.

Sense Modality	Typical Range of Reaction Time
Touch	110 ms.– 150 ms.
Audition	120 – 160
Vision	150 – 200
Temperature	150 – 200
Smell	200 – 500
Taste	200 –1100
Pain	700 –1000

The wide range of reaction times reflects differences in the speed with which the stimulation of various sensory systems leads to a motor response. The locus of such differences may be at any one of several levels or at a combination of levels. First of all, some receptors are more accessible to peripheral stimulation than others. The cold and warm receptors, for example, are located deeper down in the skin than the touch receptors. More time must necessarily elapse between the application of a thermal stimulus and the beginning of the sensory response than in the case of a touch stimulus. The substantial difference between reaction time for touch, on the one hand, and for cold and warmth, on the other, must, at least in part, be due to this difference in the accessibility of the peripheral receptors. There may be differences in the speed with which the specialized receptors themselves respond to stimulation. Finally, some sensory systems may link up with motor response systems more readily than others. We must, moreover, not neglect the possible influence of affective or emotional reactions to the various stimuli which may slow down or speed up a response. In short, the effect of the sense modality stimulated on reaction time may be either peripheral or central in nature, possibly both.

Stimulus Differences Within a Modality. *Within* a given sense modality, reaction time varies, often considerably, with the specific characteristics of the stimulus. The experimental results support the generalization that those characteristics of the stimulus which serve to increase the magnitude of response in the sense organ lead to quicker reaction times. *Intensity of the stimulus* is the first case in point. An intense stimulus produces a response of greater magnitude in the sense organ than a weak stimulus: it reaches the thresholds of a larger number of receptors, causes a larger number of nerve fibers to "fire." An intense stimulus also leads to quicker reaction time than a weak one. The relationship between intensity of stimulus and reaction time is not linear, however. As the intensity of the stimulus (say, a light or a sound) is increased above threshold, there is, at first, a rapid decrease in reaction time. Additional increases in intensity shorten the reaction time further, though less rapidly, until finally a point of diminishing returns is reached. Thus, reaction to a stimulus of medium intensity will be as fast as to one of great in-

tensity. Indeed, if the intensity becomes painful, we might expect the reaction time to be slowed up.

In the case of visual stimuli, *effective* stimulation depends not only on the intensity of the stimulus, but, within certain limits, also on its *duration* and the *size* and *location* of the area stimulated. Increases in duration of visual stimuli and in area of visual stimulation also shorten reaction time, again within certain limits. The effect of these factors is not very striking—a tenfold increase in duration may decrease reaction time by only 5 to 10 milliseconds—but the effect seems to be fairly reliable. Nor should we expect the influence of duration or area of stimulus to be very great. Visual experimentation has shown that area and duration can increase the effective intensity only over a strictly limited range of values (of intensity and time, and of intensity and duration). Finally, reaction time is shorter with foveal than with peripheral stimulation. The fovea is the rod-free area of the retina which is maximally functional in daylight, contour, and color vision. It is the area of maximal acuity for the conditions under which the visual reaction time is usually carried out (with a light-adapted retina).

The correlation between speed of reaction and effectiveness of stimulation appears also when the subject responds not to the onset (or cessation) of a stimulus but rather to a *change* or *difference* in the level of stimulation. Consider first the simple reaction time to change. The subject may look, for example, at a visual stimulus of a given brightness and is instructed to respond as soon as he perceives a change in brightness. The greater the change introduced by the experimenter, the faster will be the reaction time, even if all the changes are well above threshold. Thus, reaction time may supplement other psychophysical methods which are aimed at the measurement of a threshold and which allow the subject only a "yes" or "no" response without regard to the subjective size of the perceived difference.

The relation of reaction time to magnitude of change is graphically shown in Fig. 69. This figure also shows that the speed of response varies with the *general level of intensity* at which the subject is working. If the subject reacts to the same relative changes (i.e., equal percentages of increment and decrement) at different levels of intensity, the speed of his reactions increases with intensity. This

relationship, however, holds only up to a medium level of intensity such as shown in Fig. 69. If the general intensity level is raised further, reaction time becomes slower. Reaction time to change,

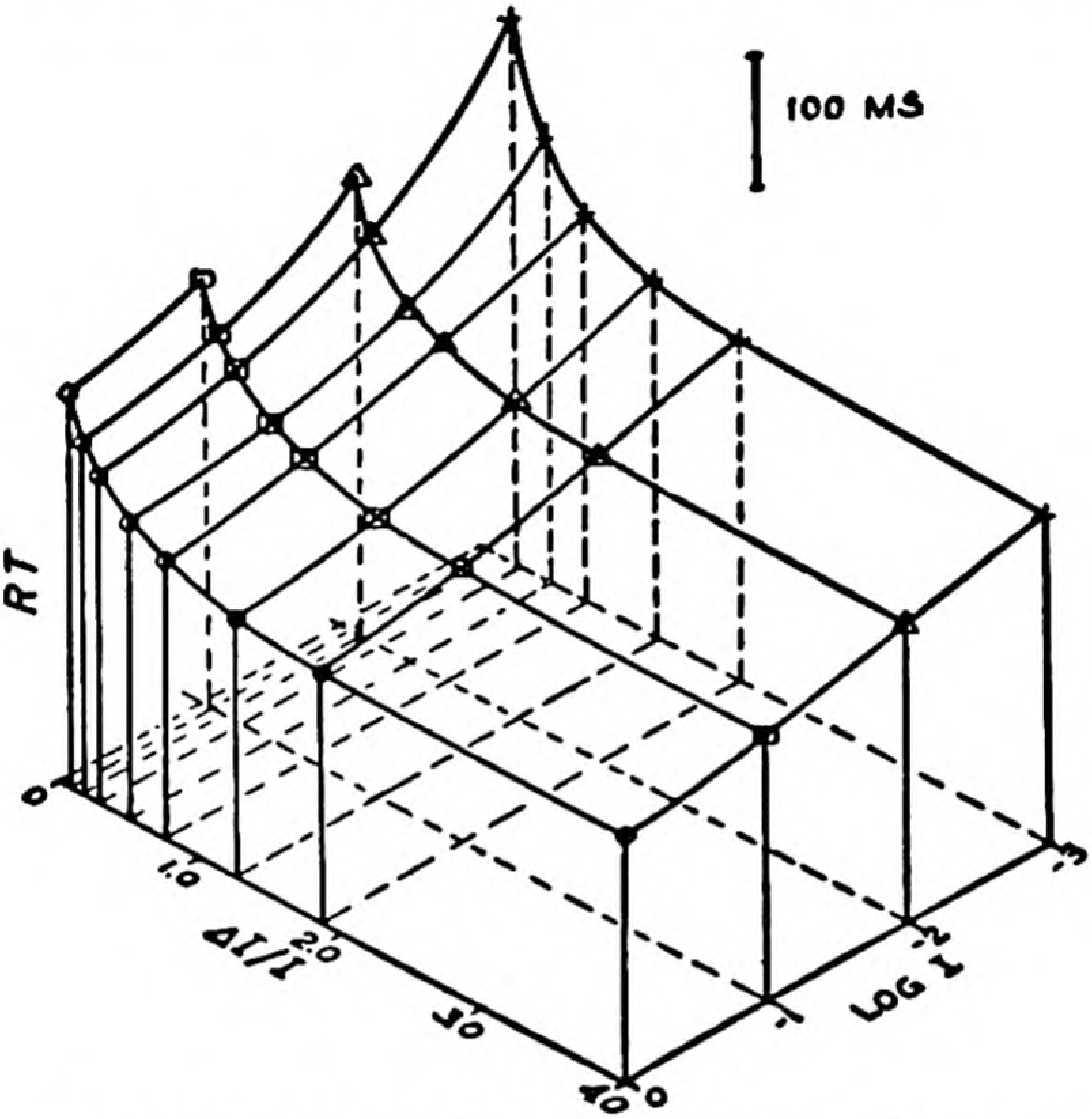


FIG. 69. The Relation of Reaction Time to Magnitude of Stimulus Change. In this graph reaction time is plotted against the relative magnitude of the stimulus change ($\triangle I/I$) at four different levels of intensity (log I). The greater the relative change, the faster is the reaction time. Within the limits shown in this figure, reaction time decreases with increases in the general level of intensity. (From A. R Steinman, Reaction time to change, *Arch. Psychol.*, 1944, No. 292, p 15, by permission of the journal and the American Psychological Association.)

then, is faster at a medium level of intensity than at a low level, but is slower at a high level than at a medium level.

Disjunctive reactions based on a discrimination by the subject yield similar results. Suppose the subject is instructed to react to the

longer of two lines presented to him, using his right or left hand according to the side on which the longer of the two lines appears. This is a reaction based on a discrimination. The larger the difference between the two lines, the quicker will be the disjunctive reaction. Again, reaction time differentiates between differences which are all well above the subject's threshold. Such results have been obtained with a variety of stimulus differences: lines of different lengths, sounds of different intensities, and colors of different hues. The way in which disjunctive reaction time varies with the magnitude of differences is closely reminiscent of a general principle of judgment discussed in Chapter 11. There, too, the speed of judgment or decision was found to be inversely related to the difficulty of the choice confronting the judge.

The dependence of reaction time on magnitude of sensory response is also supported by intersensory effects. The speed of reaction to light (or sound) accompanied by electric shock is faster than reaction time to the light, sound, or shock by itself. The addition of an electric shock, provided it is not excessively strong, may "key up" the organism and serve to facilitate the effective intensity of the other stimulus. The facilitative effect of shock in certain learning situations supports this interpretation. In summary, any condition which serves to facilitate or intensify the impact of the sensory stimulus will tend to shorten the reaction time. This generalization is well in accord with our view of reaction time as an index of the organism's readiness to respond to stimulation.

Reaction Time as a Function of Set and Attitude

Whatever the nature of the stimulus, the organism's state of *preparedness* or "set" will greatly influence the speed of response. We know well, from daily experience, that our efficiency in coping with a situation is greatest when we are fully prepared for it. The skillful hunter, lying in wait for a deer, pulls the trigger at exactly the right moment; the experienced driver waiting for a traffic signal will start his car quickly and smoothly as soon as red changes to green. The examples showing the importance of preparatory set for quick and efficient action could be multiplied indefinitely. The reaction-time experiment provides an excellent opportunity for studying the

ways in which even small fluctuations in the organism's state of readiness affect the speed and efficiency of reaction.

The Foreperiod. In the reaction-time experiment, a state of maximum readiness is produced in the subject by a "ready" signal which precedes the appearance of the stimulus by a short time. The

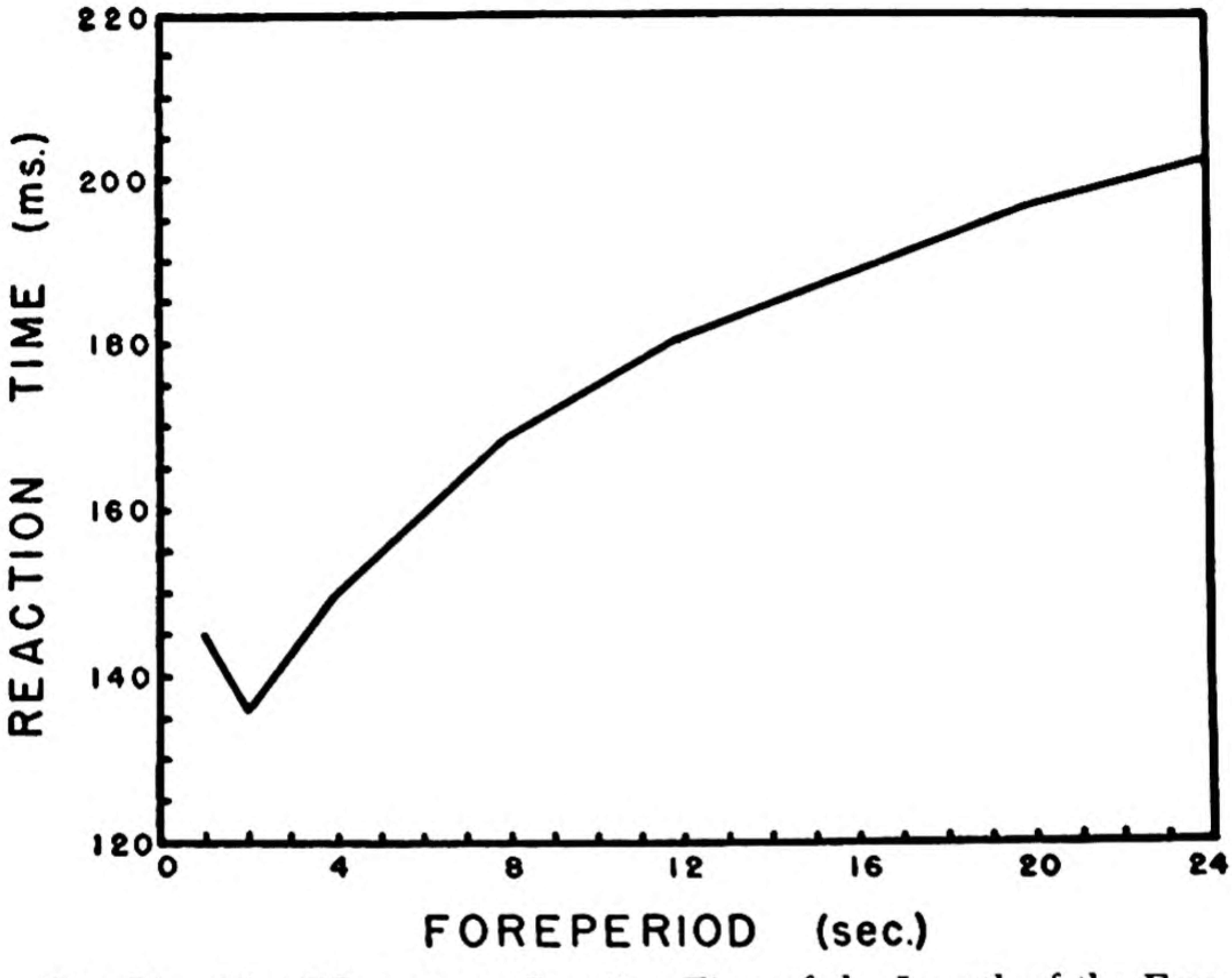


FIG. 70. The Effects upon Reaction Time of the Length of the Foreperiod. (After H. Woodrow, The measurement of attention, *Psychol. Monogr.*, 1914, *17*, No. 76, p. 26, by permission of the journal and the American Psychological Association.)

time interval between the "ready" signal and the appearance of the stimulus is known as the foreperiod.

The experimental situation raises the subject's general level of readiness. His set is one of preparedness. The "ready" signal maximizes preparedness and lowers the threshold of action. How long after the "ready" signal is the person optimally "keyed" for a response? How does reaction time vary, other things being equal, with

the length of the foreperiod? There is general agreement that a foreperiod of approximately 2 seconds is most favorable to short reaction time. Foreperiods appreciably shorter or longer than 2 seconds re-

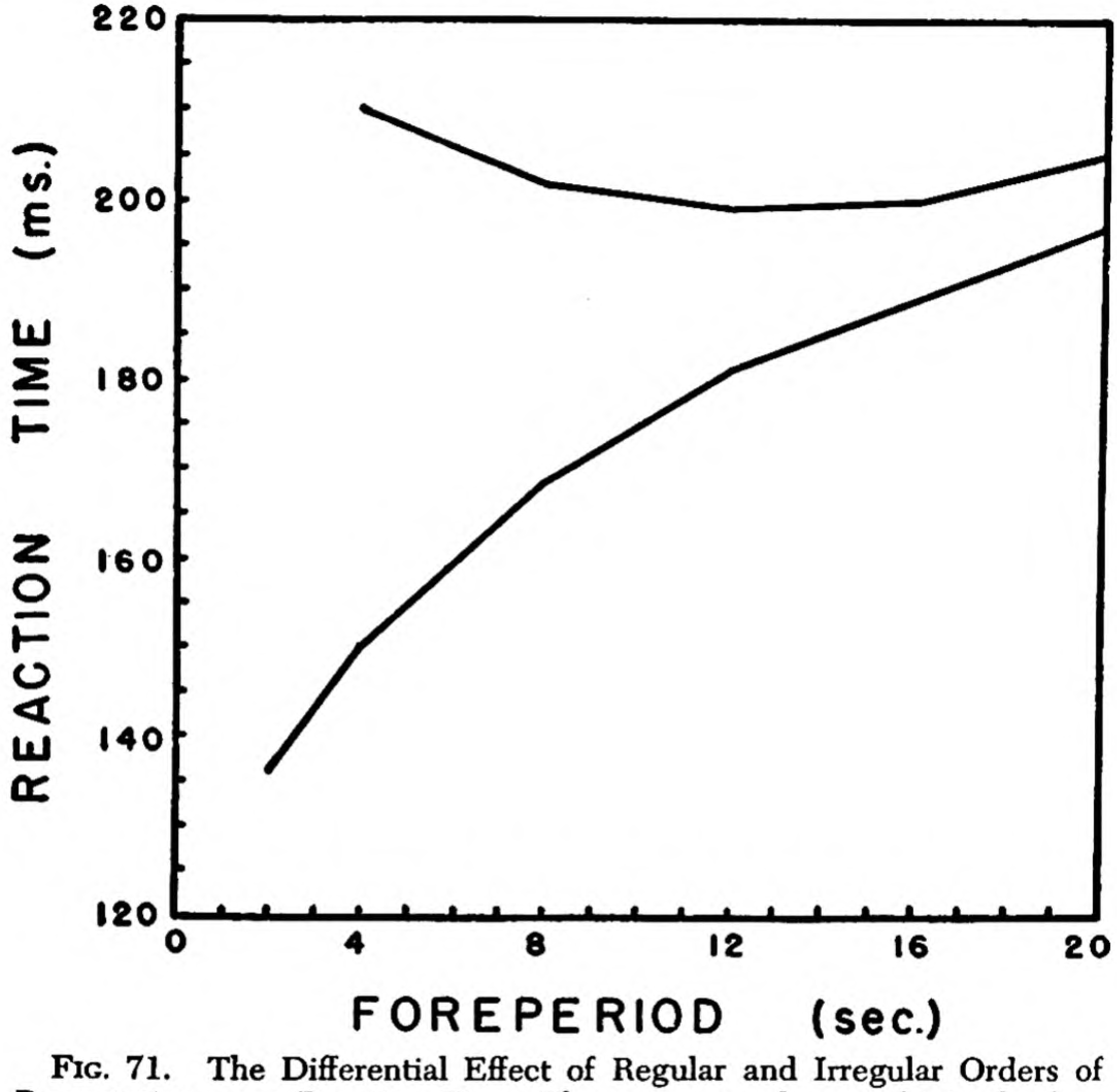


FIG. 71. The Differential Effect of Regular and Irregular Orders of Presentation upon Reaction Time. The upper graph was obtained when foreperiods were irregularly varied. The lower graph was obtained with an orderly series of foreperiods. (After H. Woodrow, Measurement of attention, *Psychol. Monogr.*, 1914, *17*, No. 76, p. 43, by permission of the journal and the American Psychological Association.)

sult in longer reaction times. Fig. 70 represents the results of a thorough experimental investigation of the effects upon reaction time to sound of the length of the foreperiod. Clearly, maximum readiness to respond is, on the average, at 2 seconds after the "ready"

signal. For individual cases, the range of optimal intervals is between 2 to 4 seconds.

The results shown in Fig. 70 were obtained when the length of the foreperiod remained constant within any block of trials and changes were introduced with full knowledge of the subject. Thus, a 1-second foreperiod would be used for twenty-five successive trials, after which the subject was informed that a 2-second period would now be used for the next twenty-five trials and so on, until the longest interval was reached. In this manner, the subject could achieve a rhythm of expectancy and adjust himself optimally to each successive interval. The results are extremely different, indeed, when foreperiods of different lengths are mixed together and presented in random order. When such an irregular sequence is used, reaction time varies very little as a function of foreperiod. The difference between a regular and irregular order of presentation is graphically illustrated in Fig. 71. For all the periods used, reaction time is longer with the irregular order of presentation. The optimal interval (and the curve does not have a very striking minimum) is displaced to the middle of the range of intervals—12 seconds—while the 2-second period now yields the longest reaction times. It seems as if the subjects adjust themselves to an average length of foreperiod and that deviations in either direction lengthen their readiness to respond.

The importance of the foreperiod emphasizes the dependence of the speed of response on the degree of the subject's readiness. Reaction time does not measure the speed with which a hypothetical "sensorimotor arc" automatically completes itself. It takes time for readiness to build up to a maximum; that is the most plausible interpretation of the 2-second minimum in Fig. 70. It is also necessary to sustain readiness over a period of waiting, or not to build up to a point of maximum readiness too quickly, if the foreperiod is too long. If the foreperiod remains the same through a series of trials, the rate at which readiness to respond is built up becomes adjusted to the duration of the period. When the foreperiod changes from trial to trial, no such adjustment can take place, and the best the subject can do is to adapt to an average length which is representative of the series and which is either too short or too long for most individual trials.

(~~PENIS~~ + ~~VAGINA~~ = CHILD)

Speed of reaction is, then, to an important extent determined by the subject's *rhythm of expectancy.* Expectancies are built up and abandoned as a result of experience. The following experiment gives a convincing demonstration of the way in which reaction time

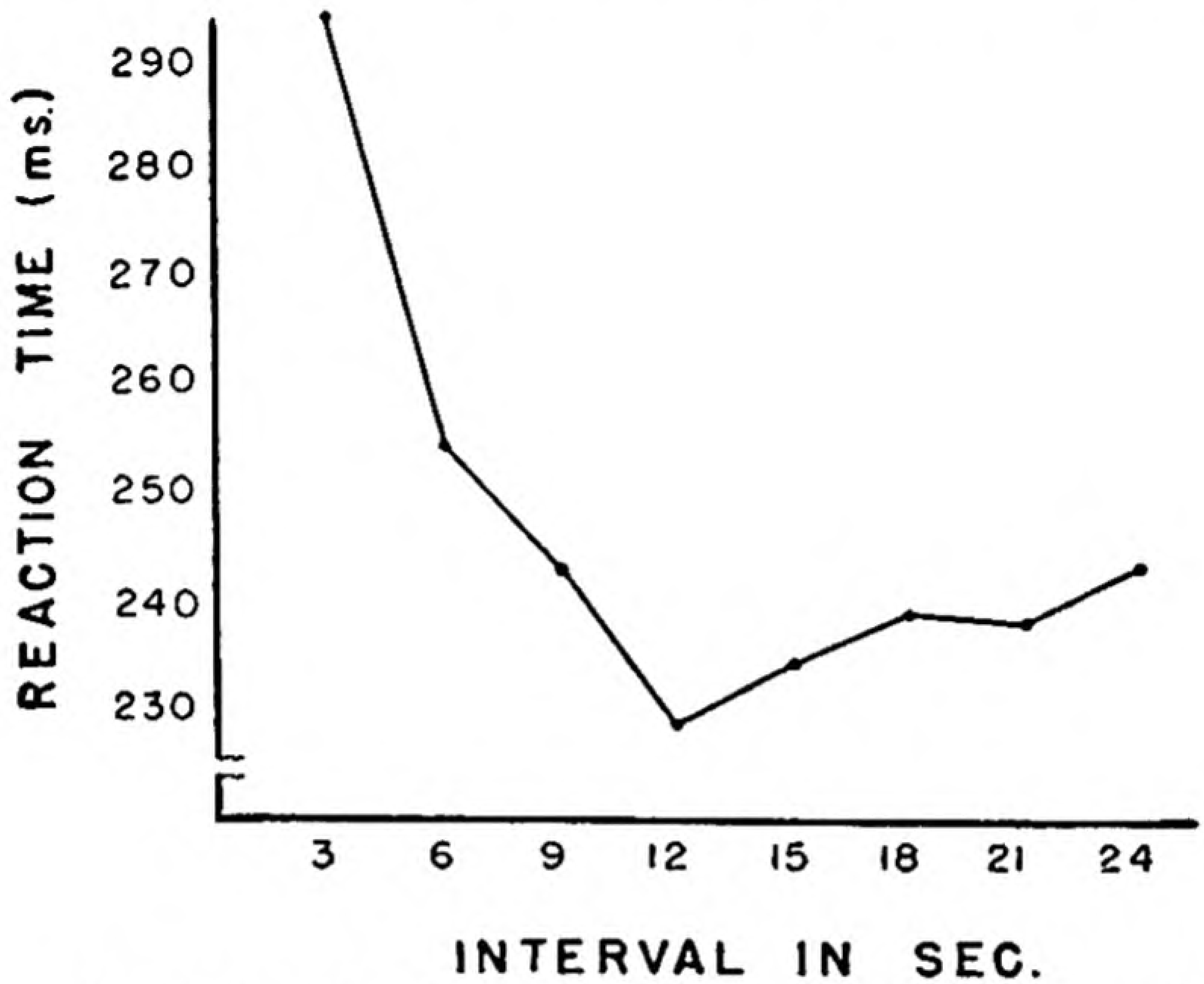


FIG. 72. The Effect of the Subject's Rhythm of Expectancy on Reaction Time. The subjects had been trained to expect the onset of a stimulus at 12-second intervals. When stimuli were interspersed at intervals other than 12 seconds, the reaction times increased as shown in the graph. (After O. H. Mowrer, Preparatory set (expectancy)—some methods of measurement, *Psychol. Monogr.*, 1940, 52, No. 233, p. 9, by permission of the journal and the American Psychological Association.)

reflects the respondent's expectancy or "preparatory set" for the stimulus. Subjects had been instructed to respond as quickly as possible to the onset of a tone delivered to them through a pair of earphones. In the first part of the experiment, these tones were presented at a fixed interval of 12 seconds between successive stimuli. The subjects built up a rhythm of expectancy and learned to respond with increasing speed. In the second part of the experiment, the 12-second interval was maintained but "test" trials were

interspersed at irregular intervals ranging from 3 to 24 seconds. As Fig. 72 shows, any interval different from the "expected" one led to an increase in reaction time. It is especially interesting to note that under these special conditions, the very short intervals led to unusually long reaction times: the subject's expectancy builds up gradually, reaches a peak at the 12-second interval. Similar curves, showing a maximum speed of reaction at the point of expectancy established through experience, could, of course, be obtained with other regular intervals. It should be further noted that not only an unexpected change in interval but also an unexpected change in the nature of the stimulus (e.g., a tone instead of a light expected by the subject) will lengthen reaction time.

Successful adjustment to a changing temporal pattern depends on the *flexibility* of the subject's expectancies, his ability to shift from one preparatory set to another. For this reason, some experimenters in the field of behavior disorders have been interested in studying the reaction times of mental patients. They found that the critical difference between normal and abnormal subjects lay not so much in the absolute values of reaction time, but in the impaired capacity of patients to adjust to long foreperiods, their inability to sustain their readiness to respond over a waiting period. This finding is of considerable theoretical interest, for it suggests that the reaction-time situation may be, within narrow limits, a paradigm for the study of broader adjustive capacities.

Sensory and Motor Attitudes. Clearly, the degree of a subject's readiness is an important determinant of his reaction time. Naturally enough, the question has been asked, readiness for what? Is it readiness for the appearance of the stimulus or readiness for the motor response? For a long time, experimenters have attempted to differentiate two attitudes of readiness or expectation in the reaction-time situation: the sensorial and the muscular. A subject with a sensorial attitude directs his attention as completely as possible toward the appearance of the sensory stimulus—the light, sound, touch, etc. A subject with a muscular attitude, on the other hand, concentrates on the motor response which he is to make as soon as the stimulus appears, e.g., the withdrawal of his fingers from a telegraph key. Usually such an attitude involves the experience of some muscular tension.

Attitudes can, of course, be only relatively sensorial or muscular. Even with concentrated attention on the stimulus, there remains a certain minimum necessary readiness for the motor response. Indeed, subjects have usually found it very difficult to maintain a sensory attitude, and the practiced performer invariably tends toward a motor type of attitude. Similarly, even an extreme concentration on the muscular response still leaves the subject in a state of preparedness for the stimulus. Sensory and motor readiness are necessarily associated. It is not surprising, then, that in the long run the distinction between sensorial and muscular attitudes has failed to yield substantial and reliable differences in reaction time. On the whole, the trend has been in the direction of shorter reaction time with a muscular than with a sensorial attitude. There have been, however, enough exceptions and conflicting interpretations to throw serious doubt on the validity of the distinction.

Range of Preparedness. Let us return to the distinction between simple and disjunctive reaction time introduced earlier. In the simple reaction-time situation, there is a single kind of stimulus and a single kind of reaction, e.g., a light and finger withdrawal. In the disjunctive situation, one of several stimuli may appear and/or one of several reactions may be required. Thus, the stimulus may be a light or a sound, the response, withdrawal of the right or left hand. The essential difference between the two situations is, therefore, one of attitude or readiness. In one case, there is readiness for a narrowly circumscribed range of events; the organism is, as it were, sharply tuned. In the other case, the set must encompass a wider range of events; the organism is less sharply tuned. When the set is thus toward a range of alternative events, we would expect the response to be less efficient, less automatic. This is indeed the case. Disjunctive reaction times are longer than simple reaction times: the difference is frequently as large as 500 milliseconds.

The wider the range of events for which the subject must be prepared, the longer becomes his reaction time to any one of these events. This fact can be demonstrated experimentally by increasing the number of alternatives in a disjunctive reaction-time experiment. There is a steady increase in the reaction time as the number of alternatives is made larger and larger. With visual stimuli, for example, simple reaction time is about 180 milliseconds. With two

alternatives, it may well exceed 300 milliseconds and when the number of alternatives is as large as ten, we may obtain a reaction time of over 600 milliseconds.

Individual Differences in Reaction Time

The reaction-time experiment is part of the mental-test tradition. When mental testing was introduced in the United States, considerable emphasis was put on the measurement of "simple" capacities. Speed of response was one of these capacities. Probably for this reason there has always been considerable interest in the problem of individual consistency and individual differences in reaction time.

Individual Consistency. Even though we speak of typical reaction times to light, sound, and other stimuli, we must remember that reaction times are highly variable. Even under constant conditions of testing, one subject's reaction times will characteristically cover a wide range of values. Some reactions will be very short, some very long. Atypically long reactions are more likely to occur than unusually short ones, i.e., the distribution of reaction times tends to be positively skewed. There is a physiological limit below which reaction time cannot be reduced, whereas there is no comparable ceiling for long reactions.

Reaction time can be considerably shortened by practice. Probably such improvement is due in large measure to a more effective preparatory attitude on the part of the subject and on a better control of the "muscular" component of the response. Obviously, practice can improve the speed of response only within narrowly circumscribed limits. Reaction time cannot be indefinitely shortened, and the physiological limit causes continued practice to yield diminishing returns.

Subjects can, then, learn how to speed up their reactions. They are aided in such improvement by the introduction of proper incentives. In one study, both positive and negative incentives were introduced in counterbalanced order and compared with the standard "neutral" situation. The positive incentive was provided by giving the subject knowledge of his performance, inducing him to set speed goals for himself and to strive for their attainment. The

negative incentive consisted of an electric shock which was delivered automatically whenever the reaction time exceeded a certain value. The introduction of an incentive appreciably shortened the reaction time as compared with the neutral situation. The negative incentive proved the more effective of the two, yielding an average improvement of about 15 percent in time as against a shortening of the reaction time by 6 percent in the presence of the positive incentive. The study also showed that the effects of an incentive on reaction time, though appreciable, are short-lived. Such effects are most noticeable in the reactions following shortly after the reward or punishment, but they quickly wear off and may even be followed by a compensatory slow-up of response. Nor are such improvements carried over from one day to the next. Incentives probably intensify, temporarily, the subject's attitude of readiness and concentration on the response, and bring about a "state of much keener attention."

Not only do reaction times of a single subject vary considerably from moment to moment, but there is also wide variation from subject to subject. Investigators have hoped to relate individual differences in reaction time to broader characteristics of the person.

Is Reaction Time Related to a General Speed Factor? Is reaction time the manifestation of a general speed factor and do extensive individual differences reflect the distribution of such a general speed factor in the population? If this were the case, we would expect high intercorrelations between (1) various reaction-time measures, and (2) between reaction time and other measures of response speed.

The evidence is not in favor of one general speed factor; nor does it point to complete independence of various speed measures. There are high correlations between simple reaction times to various types of stimuli. The correlations between performances in different reaction-time situations, such as simple and disjunctive reactions, are positive but lower. Finally, there are only very low correlations between reaction times and other performances, such as tapping and cancellation of letters, in which speed is essential. Nor are there very striking correlations between reaction time and measures of "physiological" speed such as pulse rate. Again, the correlations tend to be positive but low.

II. ASSOCIATION EXPERIMENTS

The reaction-time experiment is part and parcel of an experimental tradition which has put its main emphasis on the analysis of behavior into stimulus and response elements and which has used the concept of *association* as its primary principle of organization of behavior. In the reaction-time experiment, we are concerned primarily with the speed with which a sensory stimulus can evoke an associated motor response. In the same experimental tradition, much interest has centered on the various *kinds* of response that a given stimulus can evoke, and the differences in speed among various kinds of responses. In this section, we shall be concerned with experimental investigations of association, especially those which call for verbal responses and use words as stimuli.

The Concept of Association. A few words should be said about the concept of association which gives rise to these experiments and pervades their interpretation. Analysis of behavior into elements—however broadly or narrowly these elements may be defined—at once raises the question of the law or laws governing the *sequence* of elements. The coherent flow of language and thought, the orderliness of complex series of motor acts, in short, the reliable functioning of our vast hierarchy of habits—all these illustrate the lawful sequence of acts (stimulus and response elements) which characterizes behavior. To account for such lawful sequences, the principle of association has been invoked. Laws governing the formation of associations have been stated. Most prominent among these has been the principle of *contiguity*, according to which events that occur together or in immediate succession are associated. After *A* and *B* have occurred together or immediately following each other, the occurrence of *A* alone is likely to bring about *B*.

The concept of association has had a long and distinguished history in experimental psychology. The definition of the elements which are associated has been varied, and there has been much debate about the laws of association. In addition to contiguity, several principles—such as frequency, similarity, and vividness—

have been suggested at one time or another, but they have all remained secondary to contiguity.

The concept of association has been so hardy because it has been so useful to the experimenter in two broad areas: (1) many phenomena in the field of learning are conveniently described and analyzed as the establishment of new associations, (2) many reactions to environmental stimuli (especially verbal reactions) may be ascribed to the functioning of already well established associations. The experimental procedures about to be described have as their purpose the sampling and analysis of the rich store of verbal associations which man accumulates through the use of language.

Types of Verbal Association Experiments

In verbal association experiments, the subject's task is to respond by a word or words, usually as quickly as he can after presentation of the stimulus. According to the type of stimulus used, we can group such experiments under two general headings: *object-word* and *word-word* association.

Object-Word Association. In the object-word experiment, the stimulus may be almost any kind of object—a piece of colored paper, a geometric design, an object of daily use—and the subject is instructed to react to this object with a word, e.g., to name the object. Or, it may be a visually presented printed word which must be read aloud. Since the spoken word carries the meaning of the object, such experiments are designed to test the speed with which the verbal meaning of an object can be reported. Verbal associations are slower than simple motor reactions. A subject will, for example, take about 500 milliseconds to name a color, as compared with a motor reaction time of 180 milliseconds to a simple light stimulus. A naming response, moreover, is slower than a reading response: to read the printed name of a color requires on the average about 350 milliseconds.

In summary, then, the object-word association experiment measures the speed of verbalization. The question is: How quickly can the appropriate verbal response be given in the presence of an object or conventional symbol? The answer is that verbalization tends to be rather slow, certainly as compared with a simple motor reaction. By a simple motor reaction, the subject merely indicates

that "something is there." In giving a verbal association, the subject must make a *selective* response—the word must be appropriate to the object. Such a selective reaction takes time.

Word-Word Association: Free and Controlled. In word-word association experiments, both stimulus and response are spoken words. The relation between the stimulus word and the response word may or may not be prescribed. In the *free-association* experiment, the subject responds with the very first word that occurs to him. In the case of *controlled association*, the type of response is determined by the instructions given to the subject. Thus he may be required to respond to a stimulus word by naming its opposite (gay-sad, big-small, etc.), or by naming the genus of which the stimulus word is a species (dog-mammal, trout-fish, etc.), and so on, for any variety of relationships. Such control of association may be more or less complete. There may be only one correct response to the stimulus word (as in naming the genus of which the stimulus word is a species), or there may be several alternative responses (as in naming a species when the stimulus word is a genus).

The contrast between *free* and *controlled* association is perhaps somewhat misleading. There is no implication that one is haphazard and the other lawful. Both types of association alike must be considered to be lawfully determined. In the case of controlled associations, the experimental instructions provide a *determining tendency*—the direction of the associations is regulated by the instructions. We may say that the association is controlled externally. When association is free, the determination is, as it were, internal. The subject's past experience, his motives and habits of thought constitute the determining tendencies which cause him to respond with one word rather than another.

The Speed of Word-Word Associations. The speed of both free and controlled associations varies over a wide range, depending on the particular word in conjunction with the kind of instruction. In general, *the more fully controlled the association is by instruction, the shorter the reaction time tends to be*, provided, of course, such factors as familiarity of the words are held constant.

Since association times are distinguished by great variability from word to word and from subject to subject, it is difficult to state

truly representative average values. For controlled word association, a median value of a little less than 1 second probably represents a fair estimate. For free associations, the average value should be put higher. Averages as low as 1100 milliseconds and as high as 2000 milliseconds have been reported. The most notable fact about the distribution of free-association times is the unusual slowness of some responses. Responses which take 5 or even 10 seconds are infrequent but not out of the ordinary. (As we shall see, such very slow responses may sometimes be indicative of emotional blocks.)

Classification of Associations

The analysis of verbal associations does not end with measurements of their speed. The *kinds* of responses which are given are of interest. Investigators have hoped for a long time that the classification of verbal associations would provide an important instrument for the study of individual differences in cognitive functions. Two main approaches have been used in the evaluation of verbal associations: (1) *frequency tables*, and (2) *content analysis.*

Frequency Tables. When we administer a free-association test to a large representative sample of subjects, we obtain a distribution of responses to each stimulus word. We can tabulate the frequency with which the various responses are given, ranging from the most frequent associations given by a plurality of the subjects to those which are highly individual and given only by one in a thousand subjects. By means of such a *frequency table,* we can evaluate any given association in terms of the relative frequency with which it occurs in the sample.

To provide a standard basis for the measurement of individual differences, several sets of such frequency tables have been worked out. The first, and best known, are the *Kent-Rosanoff* frequency tables. They contain frequency distributions of responses given by a sample of 1000 subjects to 100 common nouns under standard conditions. An example of the frequency tables, listing the responses to the word *white,* appears in the following table.[1]

[1] G. H. Kent and A. J. Rosanoff, A study of association in insanity, *Amer. J. Insan.,* 1910, 67:37-96, 317-390.

Frequency of Responses to the Word *White*

1 almost	3 cotton	1 lady	1 ribbon
2 apron	1 cream	1 lawn	1 rightness
	2 curtain	1 lead	2 rose
1 baby	1 curtains	1 lemon	
1 beach		1 lie	1 sand
1 beautiful	35 dark	51 light	1 Sarah
1 beauty	1 darkness	3 linen	1 shade
1 bird	1 day	1 lovely	6 sheet
308 black	1 daylight		1 shoes
1 bleached	2 dazzling	2 man	1 shroud
9 blue	1 delicate	1 marble	1 silvery
1 boat	2 dove	9 milk	1 simple
1 Bob	34 dress	1 Mountain	1 skirt
1 body	1 dresses	1 Mountains	2 sky
1 bride		3 muslin	91 snow
4 bright	2 easy		1 snowflake
1 brightness	1 evening	1 napkin	1 snowy
1 Broadway		1 nearly	1 soft
2 brown	1 face	2 nice	1 soul
	2 feathers		1 space
2 cat	2 flag		1 spread
1 cerement	2 flower	5 paint	2 still
1 chair		2 pale	1 summer
3 chalk	2 garment	17 paper	1 sunlight
1 cheerful	1 ghost	1 pencil	1 swan
1 cherries	1 glare	2 person	
10 clean	1 good	1 pigeon	4 tablecloth
2 cleanliness	2 gray	2 pink	1 tent
1 cleanness	6 green	1 pleasing	1 tile
2 clear		1 powder	1 trees
17 cloth	1 hall	1 pretty	1 trousers
1 clothing	2 handkerchief	20 pure	
4 cloud	1 hands	19 purity	2 waist
2 clouds	1 hard		6 wall
1 coat	2 horse	1 race	1 wash
170 color	4 house	7 red	2 wedding
1 colored		1 restful	
11 colorless	1 innocence	1 retired	7 yellow

The sample used by Kent and Rosanoff included subjects of both sexes whose ages ranged from eight to over eighty years. They came from different geographical areas and varied widely in intelligence and education. Their subjects were not, however, a truly repre-

sentative sample of the population, and this fact makes the application of their frequency tables somewhat uncertain. There are other frequency tables based on more systematically controlled samples: the *O'Connor tables* obtained from a sample of 1000 adult men, and the *Woodrow-Lowell tables* based on the responses of 1000 urban school children. In using these tables, it is important to bear in mind the nature of the sample on which they are based. If the sample is not representative of the population to which the subject belongs, the use of the tables becomes, at best, hazardous. Such a procedure would be analogous to the use of test norms with subjects to whom the norms do not apply, such as the application of adult intelligence norms to the performance of children or vice versa.

Analysis of word associations by means of frequency tables has shown that subjects may differ widely in the conventionality of their responses. A measure of conformance can be obtained by finding the frequency with which a subject's responses are given in the population. The best measure of conformance is, perhaps, the median of population frequencies for the responses given by him. When this measure is applied, the most striking differences are found between normal subjects and mental patients. Whereas normal subjects give, on the average, over 90 percent common reactions, the percentage of such responses drops to about 70 in the case of patients. Schizophrenics in particular are given to bizarre, unusual associations, including neologisms and incoherent responses. By the use of frequency tables, the conventionality of a subject's responses can thus be gauged—the degree to which his associations form common or unique patterns. It is interesting to note that the more frequent a response, the more quickly it is given. This generalization is known as *Marbe's law.* Frequency and speed are two complementary ways in which the effective strength of associations expresses itself. We recall again the parallel to measures of learning. The well-established response is not only elicited frequently and regularly but also more quickly.

Content Analysis. Not satisfied with sheer frequency counts, investigators have long been interested in the classification of associations according to the content of the responses. They hoped to find significant correlations between the content of associative re-

sponses and intellectual, aesthetic, and emotional characteristics of the subjects. To this end, many different classifications of associations have been attempted. The variety of associative responses is rich, indeed, and the categories used in their classification have been correspondingly numerous. They were based on the grammatical and logical relationships between the stimulus and the response word, on the type of evaluation of the stimulus word contained in the response, on the objectivity as against the egocentricity of the response, and so on. Many of these categories were not mutually exclusive. Little purpose would be served by a detailed survey of these various classifications. Instead, we shall cite here a classification proposed by Woodworth[2] which, under four general headings, subsumes most of the more specific detailed classifications used by experimental and clinical workers. These categories follow:

1. *Definitions.* These include synonyms, e.g., justice-right, and supraordinates, e.g., sheep-animal. Such responses are especially frequent when the subject is not very familiar with the stimulus word and uses the responses to clarify its meaning.
2. *Completions or predications.* Such responses tell something about the stimulus, elaborate on it. Ocean-blue, bread-eat are examples of such responses. Children are more prone to such responses than adults, and literary individuals use them more frequently than scientifically oriented ones. Such responses seem to stem from a descriptive rather than an analytic approach.
3. *Coördinates.* The stimulus-response relationship represents a juxtaposition of terms rather than a subordination or supraordination. Stimulus-response pairs like hunger-appetite, religion-church provide illustrations. These words are juxtaposed because they refer to the same realm of activity or feeling, are examples of the same class of events, and so on. The frequently used contrast responses, such as black-white, long-short, belong to the same category. Adult subjects give more coördinate associations than children. Again, a difference between individuals with a literary and a scientific orientation has been reported, coördinate responses being more frequent among the scientists. Coördinate

[2] R. S. Woodworth, *Experimental psychology,* New York: Henry Holt and Co., 1938, pp. 348 ff.

responses are more analytical or conceptual as compared with completion responses.

4. *Valuations and personal associations.* Here we have the large group of responses which stem from a subject's unique past experience, from his valuations and attitudes, needs and emotions. Some of these responses are well described as *egocentric* because they represent a personal reaction of the subject to the stimulus word.

The scheme reproduced here has the virtue of comprehensiveness and uses clear and well-defined, but at the same time flexible, categories. Although the reader may encounter alternative classifications with valid claims for consideration, Woodworth's scheme may be invaluable to the experimenter and clinician in a first organization of the results of an association experiment.

Clinical and Diagnostic Use of Associations

We have emphasized throughout that word associations should not be regarded as haphazard but as lawfully determined by instructions and tendencies within the individual. In many cases, such determining tendencies are attributable not so much to the conventions of logic and language but rather to the personal experiences, needs, and values of the individual. For this reason, the word-association procedure has been of value to clinicians in diagnosing sources of emotional disturbance.

Complex Indicators. The assumption is made that stimulus words which tap sources of emotional disturbance (of which the subject himself may be completely unaware) will produce responses sufficiently unusual to catch the examiner's attention and to convey a suggestion of trouble. Thus, the clinicians' search has been for "complex indicators" in word association. A large number of indicators have, at one time or another, been proposed and tested but only a few have proved reliable. Perhaps foremost among these is an unusual increase in the association time. Although association times are typically quite variable, each subject has a normal range within which most of his responses fall. Response times falling outside this typical range, especially unusually long ones, frequently spell trouble. Sometimes it is not the "charged" word itself which yields the abnormally long association time but the one following it,

as if it took some time for the full impact of the emotional stimulus to manifest itself.

Other disruptions of the normal response pattern have also served as complex indicators. Among these are repetition of the stimulus word, misunderstanding of the stimulus, and highly incongruent or bizarre responses. It seems as if the subject were using such means (in all likelihood without being aware of doing so) to defend himself against the full emotional impact of the stimulus word.

There is, then, a variety of complex indicators which an emotionally charged stimulus word may arouse. How are these indicators related to each other? Does the occurrence of one increase the probability that the others will appear? If the indicators were systematically associated with each other, our confidence in their validity would be increased. A thorough investigation of the correlations among the various complex indicators has shown that a few of them do, indeed, tend to be regularly accompanied by others, notably *long association time, misunderstanding of the stimulus, and repetition of the stimulus.* When one of these indicators is present, there is a good chance that some other emotional response will appear as well. This correlation among indicators is presumptive evidence for their validity.

Diagnostic Applications. As a diagnostic tool, the word-association experiment has had two main uses: (1) as an adjunct to other methods of psychiatric diagnosis, and (2) as one in a battery of procedures aimed at the detection of guilt.

The word-association procedure may help the psychiatrist localize the general area of his patient's difficulties. Thus, C. G. Jung, the psychiatrist who pioneered in the study of word associations, devised a standard list of 100 stimulus words related to common sources of emotional difficulties (sickness, financial difficulties, marital problems, etc.). He found that the diagnostic hunches provided by the word-association test were frequently borne out by more intensive psychoanalytic procedures. Additional validation of the complex indicators is provided by correlations with other indices of emotional disturbance, such as the galvanic skin response and fluctuations in respiration. The correlations have tended to be positive but not always statistically significant.

The word-association test is used in a similar manner in the de-

tection of guilt. A word list is prepared in which words related to the crime, such as the names of the objects stolen, are scattered through a list of presumably "neutral" items. This test is administered to the suspect as well as to control subjects. If the suspect shows unusual disturbances in response to the critical words, such as significant increases in reaction time, as compared with the controls, this fact is taken as presumptive evidence of his guilt. The uncertainties of such a procedure are obvious. It is difficult to be sure that the neutral words with which the critical words are compared are indeed neutral. In addition, emotional responses to the critical words may not necessarily be due to guilt. The subject may be acquainted with the details of the crime, and aware of the suspicion resting on him. Under such circumstances he may well react emotionally to the critical words!

In spite of such limitations, the word-association technique has at times helped in the discovery of guilty persons, though failures also have been reported. The technique has received dramatic (though indirect) confirmation through the use of hypnosis. It is possible to suggest to a subject under deep hypnosis that he has committed a certain crime and, thus, to induce guilt feelings in him. As long as the hypnotic suggestion lasts, these guilt feelings may, indeed, be profound. The subject is given a word-association test before and after the hypnotic suggestion has been accepted by him. His responses to the critical crime stimuli differ significantly on the two occasions. After he has accepted the suggestion, guilt indicators appear among his responses—lengthened association times, repetitions of the stimulus, etc. When the suggestion is removed, the guilt indicators disappear. Thus, the experimental manipulation of guilt feelings produces predictable changes in the word associations. It must be remembered, however, that deep hypnosis is a very unusual state, indeed, and it would be rash to generalize too confidently from such experiments to the general validity of the word-association technique.

In summary, then, the word-association technique is at best an adjunct to other diagnostic techniques. It may help to narrow down the range of suspects, to locate the probable focus of a patient's emotional difficulties. No definite conclusion can, however, be based on word associations alone. It must always be considered in con-

junction with other techniques and carefully checked by other evidence.

Experiment XVII

Simple and Disjunctive Reaction Time

Purpose. To investigate the difference between simple and disjunctive reaction times. The specific problem is: How is reaction time influenced by the difficulty of the discrimination which must precede the response?

Apparatus. A standard reaction-time apparatus, as described on pp. 243-245 and shown in Fig. 68, is required. In this experiment, two different visual stimuli—a red light and a green light—will be used and the apparatus should be so adjusted that the experimenter can switch at will from one to the other. To this end, there should be two *pairs* of lights: a red and a green pair on the left and a similar pair on the right. Correspondingly, during part of the experiment, the subject needs two response keys: one for the right hand and one for the left. As usual, the experimenter by pressing his key delivers the stimulus and starts the clock; the subject stops the clock by withdrawing his hand from the key.

Procedure. There is an experimental group and a control group. For the experimental group, the experiment falls into two parts: (1) a simple reaction-time series, and (2) a disjunctive reaction-time series. A control group is required to take into account the effects of practice. In order to obtain an estimate of the effects of practice, the control group is given two successive simple reaction-time series. The experimental design may be summarized as follows.

	Part I	Part II
Experimental Group	Simple reaction time	Disjunctive reaction time
Control Group	Simple reaction time	Simple reaction time

The two groups should be matched according to their performance in Part I. Only if the average reaction times of the two groups are approximately the same in Part I, can the difference between the conditions in Part II be properly evaluated.

The subject is seated in front of his reaction key, separated from the experimenter and the chronoscope by a partition. The stimulus lights are mounted so as to appear comfortably in the subject's line of vision.

For the simple reaction-time series, only one of the stimulus lights, say, the red one, is used, and it is always presented in the same position.

The subject should use his preferred hand on half the trials, his non-preferred hand on the other half of the trials. This alternation between the two hands will make it possible to compare directly the results of this series with those of the disjunctive reaction-time series. At the beginning of the series, he is given the following instructions:

> "This is an experiment in reaction time. We are interested in measuring the speed with which you can react to the appearance of a light. The light will always appear in the same place (experimenter demonstrates the light stimulus). When I say 'ready,' place your index finger on the telegraph key. As soon as you have seen the light, withdraw your finger from the key. Be sure to respond as soon as you have seen the light but try never to withdraw your finger before you have seen the light."

The subject is then given twenty to thirty practice trials to acquaint him with the procedure. The results of these practice trials are not used in the final computation of the experimental results. Following the practice series, the subject is given 100 trials which constitute the simple reaction-time series. Throughout ~~this~~ series, the duration of the foreperiod—the interval between the "ready" signal and the delivery of the stimulus—is held constant at 2 seconds. Following each trial, the experimenter reads the chronoscope, records the reaction time on a prearranged score sheet, resets the chronoscope, and then proceeds to the next trial.

Sometimes the subject may give premature reactions, i.e., withdraw his finger from the response key before the stimulus has been presented. Such reactions may occur if the subject's readiness to respond is very great, causing him to "jump the gun." To reduce the number of premature reactions, it is advisable to intersperse a number of "catch trials" at irregular intervals throughout the series. In a catch trial, the "ready" signal is given as usual but is not followed by a stimulus. If a subject, nevertheless, gives a response, his error is pointed out to him, and he is cautioned to respond only after seeing the stimulus light come on.

After the simple reaction-time series, the subject is given a rest period. Following the rest interval, the members of the control group are given another hundred simple reaction-time trials. The experimental group proceeds to the disjunctive reaction-time series.

For the measurement of disjunctive reaction time, both the green and red light stimuli are used, and the subject is provided with two response keys, one for his right and one for his left hand. The following instructions are given:

> "In this part of the experiment, we shall continue to measure the

speed with which you can react to the appearance of a light. This time, however, we shall present either a green light *or* a red light.

We shall alternate between these two lights in random order. The red light will sometimes appear on the left, sometimes on the right, and the same is true of the green light. Your task is to react to the red light only, not to the green light. You are, moreover, to respond on the same side on which the red light appears. When I say, 'ready,' place the index finger of your left hand on the left key, and the index finger of the right hand on the right key. If the red light appears on the right side, withdraw the finger of your right hand; if the red light appears on the left side, withdraw the finger of your left hand. If the green light appears, do not withdraw either finger."

The experimenter then presents the red and green lights in a prearranged random series, at the same time varying the positions of the lights at random. A typical group of trials may be: *RL* (red left), *GR* (green right), *RR*, *GL*, *GL*, *RL*, *GR*, etc. As before, the foreperiod is kept constant at 2 seconds, and reaction time recorded after each trial. There is again a practice series of twenty to thirty trials followed by a series of 100 test trials.

Wrong as well as correct reactions should be noted. Wrong reactions may be (1) responses to the incorrect stimulus (the green light), (2) reactions with the wrong hand, and (3) premature reactions.

Treatment of Results. For each subject, the average reaction time in Part I and Part II is computed. We may then ask the following questions: (1) Was the average reaction time of the control subjects significantly shorter in Part II than in Part I? If so, this effect may be ascribed to practice. (2) Is there a significant difference between simple and disjunctive reaction time? To answer this question, we compare the average reaction times of the control group and the experimental group in Part II. In Part II, both groups were equally well practiced. If their reaction times are significantly different, we may ascribe this discrepancy to the difference between simple and disjunctive reaction time. We complete our analysis with a comparison of incorrect, especially premature, responses under the two conditions. If a difference is found, to what features of the two situations may it be attributed?

Experiment XVIII

Detection of Guilt Through Word Association

We shall describe the conventional experiment using the word-association technique as a method of detecting guilt. Experience has shown that such experiments are as likely as not to be failures, especially under the

artificial conditions of a classroom demonstration with laboratory-produced guilt. Nevertheless, the performance of this experiment is probably instructive in highlighting the limitations of this technique which, along with other lie-detecting devices, has long enjoyed considerable, and frequently unwarranted, popular appeal.

Purpose. To demonstrate the use of the word-association technique as a method for the detection of guilt.

Materials. A list of stimulus words containing neutral and "critical," i.e., guilt-relevant, items is constructed. The critical words should refer to things unequivocally associated with the "crime." The neutral words are neutral in the sense that they refer to items which are in no recognizable way associated with the crime. Whether or not they may have emotional connotations for the subjects due to other experiences is unknown. A voice-key arrangement provides the most precise method for measuring the association time (see p. 245). If voice keys are not available, a stop watch may be used for good approximations to the association times.

Procedure. This experiment is best performed as a classroom demonstration. The experimenter chooses a small group of subjects (three is an adequate number). All the subjects are sent out of the room. One has been privately instructed to commit a "crime," the others serve as controls. The controls, of course, are kept in ignorance of the nature of the crime. The nature of the crime must be left to the ingenuity of the experimenter in using available facilities. Sometimes it is possible to find a coöperative librarian who agrees to the theft of a book. Another favorite misdeed is the killing of a rat. Again, other experimenters have required their subjects to read highly emotional passages. Whatever the crime, the critical stimulus words refer in detail to its nature and to the circumstances under which it was committed.

While the subjects are out of the room, the procedure is fully explained to the class and the neutral and critical items in the word list are identified. The audience is also reminded of the reactions which are commonly regarded as complex indicators: long association times, repetitions of the stimulus word, misunderstanding, laughing, blushing, etc.

The crime having been committed, the culprit and the controls submit to the word-association test. They are, of course, tested one at a time. The subject is seated comfortably, preferably with his back to the experimenter whose facial expression may otherwise influence the responses. The subject is instructed to respond to each stimulus word with the first word that comes to mind, regardless of any logical or grammatical relationship. The speed of the associative responses is read from the chronoscope (if voice keys are used) or from the stop watch. The re-

sponse words are written down as well as any relevant characteristics of the subject's behavior, such as laughing, blushing, and other signs of embarrassment.

It is possible to combine the word-association test with other measures of emotional response, such as the galvanic skin response and changes in the breathing cycle. For a description of these techniques, see Chapter 19, Experiment XXVII.

Treatment of Results. The main quantitative data are, of course, the association times. The average association times for neutral and critical words are computed for each of the subjects. In each case, the size and reliability of the difference in association time for neutral and critical words are determined. This analysis is supplemented with evaluation of the subjects' behavior during the test—the extent to which they showed signs of guilt in the presence of the critical words. If physiological indices of emotion have been obtained, these are, of course, added to the picture. On the basis of all the available evidence, the members of the class are asked to indicate the most likely suspect. The vote should be taken in writing to eliminate the possible influence of suggestion. In all probability the vote will be divided, perhaps the majority will condemn one of the innocent controls—an object lesson in the cautious interpretation of evidence and the pitfalls of lie detection.

General References—Reaction Time

Garrett, H. E., *Great experiments in psychology,* New York: D. Appleton-Century Co., 1941, Chap. 14.

Woodworth, R. S., *Experimental psychology,* New York: Henry Holt and Co., 1938, Chap. 14.

Determinants of Reaction Time

Reaction Time as a Function of Stimulus Characteristics

Froeberg, S., The relation between the magnitude of stimulus and the time of reaction, *Arch. Psychol.,* 1907, No. 8.

Henmon, V. A. C., Professor Cattell's work on reaction time. In The psychological researches of James McKeen Cattell, *Arch. Psychol.,* 1914, No. 30.

Johanson, A. M., The influence of incentive and punishment upon reaction time, *Arch. Psychol.,* 1922, No. 54.

Poffenberger, A. T., Reaction time to retinal stimulation, *Arch. Psychol.,* 1912, No. 23.

Steinman, A. R., Reaction time to change, *Arch. Psychol.*, 1944, No. 292.

Todd, J. W., Reaction to multiple stimuli, *Arch. Psychol.*, 1912, No. 25.

Reaction Time as a Function of Set and Attitude

Breitweiser, J. V., Attention and movement in reaction time, *Arch. Psychol.*, 1911, No. 18.

Mowrer, O. H., Preparatory set (expectancy)—some methods of measurement, *Psychol. Monogr.*, 1940, 52, No. 233.

Rodnick, E. H., and Shakow, D., Set in the schizophrenic as measured by a composite reaction time index, *Amer. J. Psychiat.*, 1940, 97:214-225.

Woodrow, H., The measurement of attention, *Psychol. Monogr.*, 1914, *17*, No. 76.

Individual Differences

Johanson, A. M., The influence of incentive and punishment on reaction-time, *Arch. Psychol.*, 1922, No. 54.

Lanier, L. H., The interrelations of speed of reaction measurements, *J. Exper. Psychol.*, 1934, *17*:371-399.

Lemmon, V. W., The relation of reaction time to measures of intelligence, memory, and learning, *Arch. Psychol.*, 1927, No. 94.

Miles, W. R., Age and human ability, *Psychol. Rev.*, 1933, *40*:99-123.

General References—Association Experiments

Jung, C. G., *Studies in word association. Experiments in the diagnosis of psychopathological conditions carried out at the psychiatric clinic of the University of Zuerich*, New York: Moffat, Yard, 1919.

Woodworth, R. S., *Experimental psychology*, New York: Henry Holt and Co., 1938, Chap. 15.

Types of Verbal Association Experiments

Cason, H., and Cason, E. B., Association tendencies and learning ability, *J. Exper. Psychol.*, 1925, *8*:167-189.

Cattell, J. McK., Experiments on the association of ideas, *Mind*, 1887, *12*:68-74.

Lund, F. H., The role of practice in speed of association, *J. Exper. Psychol.*, 1927, *10*:424-433.

May, M. A., The mechanism of controlled association, *Arch. Psychol.*, 1917, No. 39.

Murphy, G., An experimental study of literary vs. scientific types, *Amer. J. Psychol.*, 1917, *28*:238-262.

Woodworth, R. S., and Wells, F. L., Association tests, being a part of the

report of the committee of the American Psychological Association on standardizing procedure in experimental tests, *Psychol. Monogr.*, 1911, *13*, No. 57.

Classification of Associations

Kent, G. H., and Rosanoff, A. J., A study of association in insanity, *Amer. J. Insan.*, 1910, *67*:36-96; 317-390.

Murphy, G., An experimental study of literary *vs.* scientific types, *Amer. J. Psychol.*, 1917, *28*:238-262.

O'Connor, J., *Born that way*, Baltimore: The Williams and Wilkins Co., 1928.

Wells, F. L., A preliminary note on the categories of association reactions, *Psychol. Rev.*, 1911, *18*:229-233.

Woodrow, H., and Lowell, F., Children's association frequency tables, *Psychol. Monogr.*, 1916, *22*, No. 97.

Clinical and Diagnostic Uses of Association Experiments

Crane, H. W., A study in association reaction and reaction time, *Psychol. Monogr.*, 1915, *18*, No. 80.

Crosland, H. R., The psychological methods of word-association and reaction-time as tests of deception, *Univ. Ore. Publ., Psychol. Ser.*, 1929, No. 1.

Hull, C. L., and Lugoff, L. S., Complex signs in diagnostic free association, *J. Exper. Psychol.*, 1921, *4*:111-136.

Huston, P. E., Shakow, D., and Erickson, M. H., A study of hypnotically induced complexes by means of the Luria technique, *J. Gen. Psychol.*, 1934, *11*:65-97.

Jung, C. G., *Studies in word association. Experiments in the diagnosis of psychopathological conditions carried out at the psychiatric clinic of the University of Zuerich*, New York: Moffat, Yard, 1919.

Wells, F. L., *Mental tests in clinical practice*, New York: Wald Book Co., 1927.

Jur Jung Huston Jung, C. G., Complex Signs

»»» 13 «««

MEASUREMENT OF LEARNING

ONE of the outstanding characteristics of the behavior of organisms is the continual modification of responses through learning. From birth on, and even before, organisms acquire new responses and adapt their old ones to changing circumstances, achieving an ever-growing repertory of reactions. These habits persist in time. Some last only for a short period, while others function again and again and may do so throughout the life of an individual. *Learning* and *retention* are key concepts in the analysis of behavior. Try to imagine an organism without the capacity to learn and to retain what it has learned! Such an organism would be forever doomed to repeat the responses with which it was natively equipped, without benefit from experience and teaching, reward and punishment.

The influence of learning is universal. If we consider any area of behavior—motor and verbal skills, problem solving, attitudes and values—we find that they all reflect past learning in countless ways and are continually modified by new learning. Not only is learning ubiquitous, but it manifests itself in many forms of behavior changes and at many levels of complexity. Because of this variety and complexity of learning, it is necessary to study it analytically under controlled laboratory conditions. Through experimental analysis, some of the basic variables and relationships of learning have been isolated. With this knowledge and insight gained in the laboratory, it has been possible to apply principles of learning to many types of behavior, from simple motor performances to the socialization of the child in a specific culture.

Definition of Basic Terms

Learning. We all know what we mean by *learning*, but the experimenter does well to have a rigorous definition, especially for

terms which common usage has endowed with a rich variety of meanings. Wherever possible, definitions should be anchored in the experimental procedures from which they are derived. In defining *learning*, psychologists refer to *measurable changes in behavior as a result of practice* and the conditions that accompany practice.[1] To give one simple example: as a subject reads a prose passage over and over, he can reproduce more and more words of the passage in their proper sequence. Obviously, this progressive improvement can be quantified and related to the number of readings and other relevant conditions.

Retention. An individual who has learned to repeat a prose passage without error usually can reproduce at least part of it again at a later date. The amount that he can reproduce is a measure of his retention. By *retention*, then, we mean the *persistence of the measurable changes in behavior that have been acquired through practice*. Since the learning of most activities is a gradual process requiring a certain amount of time to reach a given level, the progress of learning implies the retention of that previously acquired. For example, our subject may learn the first two lines of the prose passage on the first reading. On the second reading, he may add another line or two. In order to reach his goal of perfect reproduction, he must retain on later trials what he has learned on the earlier ones. In the laboratory, the distinction between learning and retention refers to the time at which the measurements are made. It is customary to designate the period during which the process of acquisition goes on as the *learning period*. If a measurement of performance is made after the end of the learning period, we refer to it as a measure of *retention*.

Stimuli and Responses. We have emphasized the necessity of studying the learning process analytically. The first step in the analysis is to specify the responses which are acquired and modified and the stimulus conditions under which this acquisition and modification occur. The goal of a learning experiment is to establish lawful relationships between such conditions as frequency of stimu-

[1] It will be realized that this definition does not differentiate between what we ordinarily call learning and fatigue. Learning is inferred from certain changes in the responses made to a learning situation (usually improved performance), and fatigue from other kinds of behavior changes (usually deterioration of performance).

lus presentation, their rewarding and punishing consequences, etc., on the one hand, and the occurrence of responses on the other. Yet, clearly, it is never possible to describe *all* the possible properties of a stimulus or a response or to repeat either stimulus or response identically. As we have pointed out in Chapter 1, all we can do is to establish *classes* of events with certain common critical properties. These classes we then describe and treat experimentally as stimuli and responses. The subject learning the prose passage may read it on successive practice trials under slight changes of illumination, with different eye movements, and at slightly different speeds. On the response side, he may reproduce it again at slightly different rates, at different voice levels, and with various inflections. We do not hesitate, however, to call the words on the printed page the stimuli and the subject's oral reproduction of these words, the responses.

The degree of restriction on the classes of stimuli and classes of responses is dictated by the specific purpose of the experiment. If we were interested in determining the number of trials required for perfect repetition of the passage, clearly, it would not suffice to consider as a response a statement of the meaning of the selection. Our unit in that case would need to be the individual word. On the other hand, if we were interested in determining the number of trials necessary for the subject to reproduce the passage at a uniform rate in a perfect monotone, we would require not only that the individual words and their sequence be correct, but, in addition, that words be spoken at a uniform rate and at the same voice level. This definition of response would be a much narrower one than the definitions adopted in the previous examples. There is no one degree of restriction which is inherently superior to others: it all depends on the question asked by the experiment. In the field of verbal learning, the correct reproduction of the stimulus items, without regard to such variables as inflection and loudness of voice, is usually considered adequate.

Association. Whenever a stimulus situation comes to evoke a response which it did not evoke previously, we say that an *association* has been formed between the stimulus and the response.

The concept of *association* has had a stormy history because for a while it was used synonymously with particular kinds of asso-

ciation. For example, at one time it denoted the connection of one idea with another idea; at another time, the formation of synaptic bonds in the nervous system. The term *association*, however, has much higher generality. The fundamental aspect of association is the fact that a given stimulus acquires the property of leading to a given response. As McGeoch said, the events associated "may be as analytically small as a pinpoint of discriminated light and a single muscle twitch thereto; they may be as large as the total perceived field of objects and the most complex possible response. . . ."[2] In this sense, association is a basic concept in learning. What units are profitably singled out for the study of the process of association is an altogether different problem.

Types of Learning

Almost any kind of behavior can be modified by learning. Because of this very universality, experimental analysis is greatly aided by a classification of the various areas of behavior in which responses are acquired and modified by learning. Of course, any such classification will, to a considerable extent, be arbitrary. The behavior of organisms does not readily lend itself to classification in terms of mutually exclusive categories. The justification for a breakdown into different types of learning lies purely in experimental expediency. Different areas of behavior pose their own problems of experimental design, instrumentation, and quantification. For example, the conditioning of hand withdrawal to an electric shock obviously requires different methods of measurement than those needed in the case of learning a prose passage. For these reasons, experimenters have classified forms of behavior and materials of learning into different categories.

Verbal Learning. Man is primarily a verbal learner. When a subject enters the laboratory, he already possesses a rich repertory of verbal responses which is well overlearned and highly organized. For this reason, it has been difficult to find standardized materials for the study of verbal learning. Faced with verbal material, each subject has his own complex of associations, meanings, and preferences in relation to that material. So great did this difficulty appear

[2] J. A. McGeoch, *The psychology of human learning*, New York: Longmans, Green and Co., 1942, p. 25.

to the first experimenter in verbal learning (Hermann Ebbinghaus) that he attempted to overcome it by creating the *nonsense syllable.* The nonsense syllable consists of a vowel between two consonants and has no dictionary meaning. Examples of nonsense syllables are VUJ, WOM. The nonsense syllable has been widely used in the hope that different learners would find the materials of equal difficulty and that different conditions of learning could be directly compared.

Of course, the individual nonsense syllable is anything but nonsensical in the full sense of the word. Furthermore, even where there are no associations due to experience and similarity to conventional words, the learner soon creates them. However, a particular sequence of nonsense syllables is new to the subject, and it is relatively easy to construct many lists of syllables of the same difficulty.

There have been attempts to quantify the degree of association possessed by a given nonsense syllable. One experimenter presented nonsense syllables individually and asked his subjects to report whether or not the syllable evoked any association. Depending on the percentage of subjects reporting an association, he classified the syllables with respect to *association value.*[3] This is, of course, only one way for appraising the association value of a nonsense syllable. Another investigator had as his aim the measurement of association value in terms of actual speed of learning and, on the basis of his results, published standardized lists of syllables.[4] Helpful as these standardizations have been, there still remain wide individual differences among subjects with respect to the association values of syllable lists.

The investigation of verbal learning has not been confined to nonsense materials. Eager to come as close as possible to practical conditions of learning, investigators have not hesitated to employ meaningful words, poems, and passages of prose. The use of such material makes it difficult to compare different experimental conditions with one another because of the unequal difficulty of the learning tasks. Nevertheless, with the aid of careful controls, some

[3] A set of standardized lists obtained in this study appears in J. A. Glaze, The association value of nonsense syllables, *J. Genet. Psychol.*, 1928, 35:255-267.

[4] The lists derived from this investigation may be found in the following reference: C. L. Hull, The meaningfulness of 320 selected nonsense syllables, *Amer. J. Psychol.*, 1933, 45:730-734.

relationships of a high degree of generality have been discovered through the use of such verbal materials. For example, the use of large numbers of subjects under each experimental condition makes it likely that individual differences in learning difficulty will cancel out to some extent and not influence the results unduly.

The examples listed above by no means exhaust the sources of verbal materials. Many kinds and combinations of materials have been tried out and found useful. Thus, the consonant syllable, e.g., CKN or JFH, has been employed along with the nonsense syllable. Lists of numbers and word-number combinations have also been used.

Motor Skills. Though man may use words in practically any task he attempts, there are certain kinds of activities where words play a less important role than in the type of learning we have discussed above. This is generally true for motor skills. Take the example of an individual who is learning to play tennis. Certainly words will help him by the specification of some of the movements required for the mastery of the task. On the other hand, the execution of a smooth, expert movement does not depend on a proper verbal sequence. After sufficient practice, the verbal component virtually drops out. The study of the acquisition of such acts of skill, of course, involves somewhat different concepts and measures from those used in the investigation of verbal learning. In the measurement of motor learning, for example, great emphasis may be placed on the exact nature and speed of movements as well as on the results achieved by them. Thus, the tennis teacher may evaluate a pupil in terms of such things as his stance during a backhand stroke or in terms of his ability to compete successfully with a certain class of players.

Problem Solving. In an experimental situation, a subject may either be given a series of items and required to learn them or he may have the task of discovering which ones of many possible responses are correct. In *rote learning*, all responses are indicated, and it is the individual's goal to reproduce them as faithfully as he can. In *problem solving*, on the other hand, the subject must discover and fixate the correct response as well as eliminate the erroneous ones. Usually the problem is not solved without a certain amount of trial and error. Much of this trial and error consists of verbal re-

sponses which may either be given overtly, or their presence may be inferred from the behavior of the subject. Certainly verbal processes play an important role in human problem solving.

Here is an example of an experiment involving both the discovery and fixation of the correct response by the subject. Different Chinese characters were presented over and over, and subjects were required to associate a nonsense name with each of the characters. The characters were complex and were changed from reading to reading. However, each character had a basic component or radical, and all characters with the same radical had the same nonsense name. In this situation, learning depended on the realization of the principle according to which the association of character and name is made. The subjects did learn the correct associates. One important finding was that in some cases subjects could respond correctly without being able to verbalize the principle on which their responses were based. This experiment is a good example of human problem solving because (1) the correct responses must be discovered by the subject, and (2) verbal processes aid in the solution and learning to varying extents.[5]

Of course, in many learning situations, not only in those which involve problem solving, the subject must discover some aspects of the correct response for himself. In the case of motor skills, for example, the goal is stated and perhaps the general mode of attack is indicated to the subject. He must, however, find for himself the particular way of achieving the smooth, integrated act.

Measurement of Learning

Let us recall our definition of association: whenever a stimulus situation comes to evoke a response which it did not previously evoke, an association has been formed between the stimulus and the response. Such an association is not an all-or-none affair. One of the great achievements of the experimental psychology of learning has been to devise methods for the measurement of the strength of an association. Strength cannot be inferred from a single act of behavior. Rather, it is inferred from the quantitative characteristics of

[5] C. L. Hull, Quantitative aspects of the evolution of concepts, *Psychol. Monogr.*, 1920, 28, No. 123.

a series of responses. Let us now consider the principal methods for the measurement of strength of association.

Frequency of Response Evocation. A strong association is inferred from the fact that a response is given on nearly every occasion that the stimulus is presented. By this measure, a scale of strength is established in terms of the proportion of the total number of stimulus presentations which evoke the response. Complete absence of association is defined by 0 percent evocation; full strength of association, by 100 percent evocation; 100 percent evocation, however, does not necessarily denote the greatest possible strength of association. Thus, two associations may both yield 100 percent responses but turn out to be of different strength when tested by more sensitive indices. Moreover, the meaning of any proportion varies with the number of trials on which the proportion is based. Thus, the associations established by learning a list of nonsense syllables to a criterion of two perfect repetitions will not be as strong as the associations resulting from learning the same list to a criterion of five perfect repetitions. In general, the larger the number of trials yielding 100 percent response evocation, the greater the inferred strength of association.

Resistance to Forgetting. As we have seen, percent response evocation is not an unequivocal measure of strength. This frequency measure may be made more sensitive by introducing varying time intervals between the end of practice and the measurement of associative strength. Differences not apparent at the end of practice may be measurable after a time interval. Thus, in our previous example, the lists learned to either two or five perfect repetitions were both at 100 percent strength at the end of practice. Differences in their strength would become apparent on a test a week after the end of practice. At that time, more items would be recalled from the list learned to five perfect repetitions. In this way, resistance to forgetting provides a measure supplementary to frequency of response evocation.

Ease of Relearning When a time interval intervenes between the end of practice and the test, the associations are weakened by processes occurring during that time interval. After a given interval of time, an initially strong association is still stronger than an initially weak one. Clearly, a greater amount of practice will be

required for the weaker association than for the stronger one if the two are to be brought back to the same strength. Hence, the number of trials for relearning to a given criterion provides another measure of the relative strength of associations. Let us return to our example of lists learned to two and five perfect repetitions. If both lists are relearned after one week, the list originally learned to a criterion of two repetitions will require more trials to yield 100 percent evocation again than will the other list.

Latency. One rather different measure of the strength of an association is that of the latency of the response. Latency refers to the amount of time which elapses between the onset of the stimulus and the beginning of the response. A strong association implies a high degree of readiness to give the response under appropriate stimulus conditions. Experimental investigations have, indeed, shown that the more frequently evoked responses are also the quicker ones. Sometimes latency measures will detect differences which are not easily measured by the frequency method.

The specific experimental procedures by which these measures of associative strength are made will be elaborated in subsequent chapters.

Learning Curves

Having defined a scale of associative strength, we can proceed to plot it as a function of number of practice trials. Such a plot is called a *learning curve*. Strength of learning can be measured in various ways, and the particular units employed will help to determine the characteristics of the learning curve.

When the performance measure is the number of responses acquired, we obtain a rising curve. When the performance measure refers to the elimination of wrong responses or the time required for the completion of the task, we obtain a falling curve. For example, when a maze is learned, the number of correct choices per trial as a function of practice would show a rising curve. On the other hand, either the number of errors or the time per trial as a function of practice would yield a falling curve.

Whether rising or falling, learning curves may develop differently in time. The amount of change in performance between two successive measurements may keep increasing or keep decreasing over

a certain number of trials. If each successive change is greater than the preceding one, the resulting curve shows *positive acceleration.* Such a curve is the one labeled 2 in Fig. 73. On the other hand, if each successive change is smaller than the preceding one, the curve shows *negative acceleration* (Curve 1 in Fig. 73). Some curves show an initial period of positive acceleration followed by a period of negative acceleration. Such curves are designated as *S-shaped* (Curve 3 in Fig. 73).

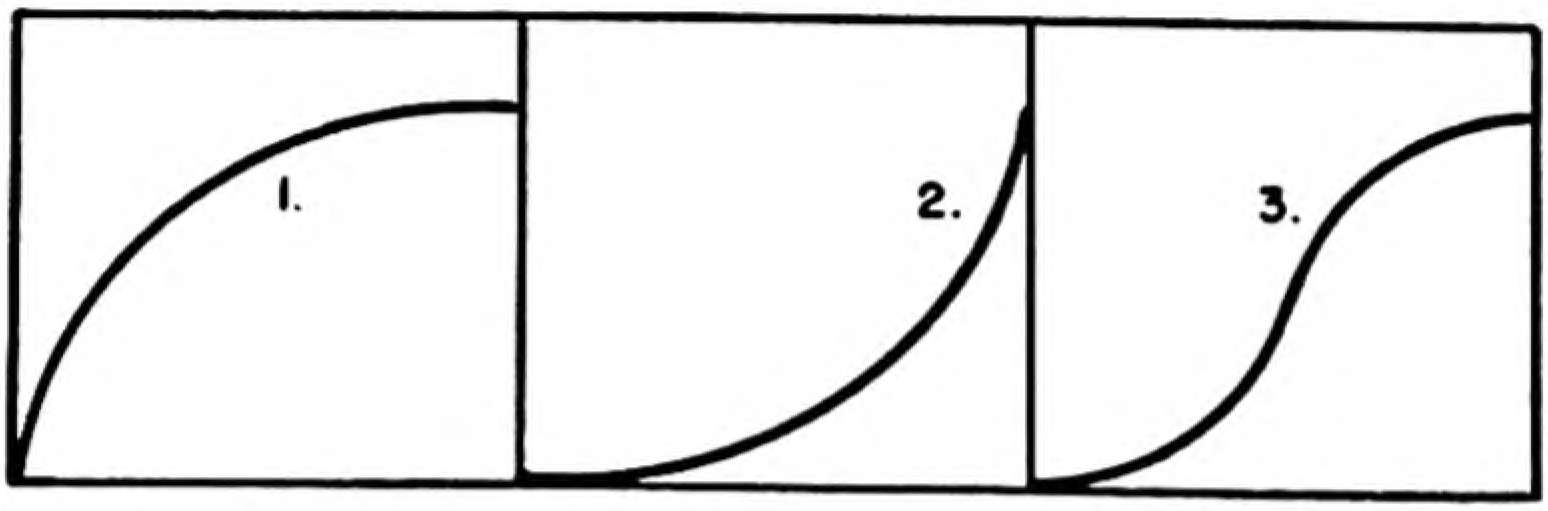


FIG. 73. Typical Learning Curves. Curve 1 shows negative acceleration; Curve 2 shows positive acceleration; Curve 3 shows an initial period of positive acceleration followed by negative acceleration (S-shaped).

There is no simple rule for predicting the form of a learning curve in advance. Certain general relations between learning procedure and the form of the resulting curve have, however, been established. For example, if the material consists of some parts which are easy and some parts which are difficult to learn, a negatively accelerated curve is typically obtained. The easier parts are quickly learned at the beginning, with the harder parts being acquired more slowly later on. Positively accelerated curves, on the other hand, are typically obtained for learning tasks which become progressively easier as practice continues. Since the S-shaped curve is a combination of the two other types, a proper combination of factors leading first to positive and then to negative acceleration yields such a curve. Even though such general relationships hold up fairly well, the particular shape of a learning curve is always jointly determined by the particular materials, the characteristics of the learner, the learning procedure, and the methods and units of measurement.

Vincent Curves. A learning curve describing the performance of a single subject is usually quite irregular, probably because it is

not possible to control adequately many of the factors which contribute to fluctuations in performance from moment to moment. For this reason, it is necessary to average the performances of many learners in order to obtain a good picture of the general relationship. The averaging of learning data, however, poses a special problem. Different individuals enter the learning situation with different capacities to learn and with different preparations for the task. Therefore, different states of acquisition will be represented by different subjects' performances on any given trial. Suppose a group of subjects has the task of learning a list of ten words. Some are fast learners, others are slow learners. On a given trial, say the fifth, the fast learners will have nearly achieved the goal of a perfect repetition while the slow learners will have learned only a few items. To average the raw data trial by trial in order to obtain a composite curve would introduce a serious distortion. To do so would mean to combine the data for individuals at different stages of learning.

One method, which at least partially takes this difficulty into account, is the construction of a *Vincent curve*. This method consists of dividing each individual's total learning period into a given number of divisions. The number of divisions is the same for all subjects, but the size of these divisions will vary from subject to subject depending on the total length of their learning periods. The performance at the end of corresponding periods are averaged to obtain the composite learning curve.

The construction of a Vincent curve may be illustrated by the following simple example. There are two subjects learning a list of 10 words. One of the subjects is a fast learner, the other a slow one. The first subject needs 5 trials to reach the criterion of one perfect repetition, the second, 10 trials. The accompanying figures show the number of correct responses given by the two subjects on each successive trial.

	Trials									
	1	2	3	4	5	6	7	8	9	10
Subject A:	2	4	7	8	10					
Subject B:	2	3	3	5	7	6	7	8	8	10

We divide the learning periods of the two subjects into five divisions each. For Subject A, each division consists of 1 trial; for Subject B, 2 trials. At the end of the first fifth of his learning period, Subject A

gives 2 correct responses, Subject B, 3 correct responses. The average value at the end of the first fifth is, therefore, 2.5. In a similar manner, the values in the accompanying table are obtained.

	Successive Fifths of Learning Period				
	1	2	3	4	5
Subject A:	2	4	7	8	10
Subject B:	3	5	6	8	10
Average:	2.5	4.5	6.5	8	10

When these average values are plotted, a composite learning curve is obtained. When the values for a large number of subjects are so averaged, a fairly smooth curve usually results.

There are certain dangers in the use of Vincent curves which should be mentioned. Such average curves should be computed only if there is reasonable assurance that each subject begins the task at the same level of performance. Thus, if one subject enters the learning situation with one-third of the learning already accomplished while another subject starts from the beginning, we are averaging together different segments of the total learning curves. We assume when we use the Vincent method that each subject would give the same *form* of the curve even though some subjects take more trials than others. Unless all subjects start at about the same level, the method is likely to be misleading.

General References

Hunter, W. S., Learning: IV. Experimental studies of learning. In C. Murchison (ed.), *A handbook of general experimental psychology,* Worcester: Clark University Press, 1934.

McGeoch, J. A., *The psychology of human learning,* New York: Longmans, Green and Co., 1942, Chaps. 1 and 2.

Melton, A. W., Article on learning. In W. S. Munroe (ed.), *Encyclopedia of educational research,* New York: The Macmillan Co., 1941.

Robinson, E. S., *Association theory today: an essay in systematic psychology,* New York: Century, 1932.

Skinner, B. F., *The behavior of organisms,* New York: D. Appleton-Century Co., 1938.

»»» 14 «««

CONDITIONING

THE study of conditioning was begun by physiologists and then enthusiastically adopted by psychologists who saw in it an avenue to the objective investigation of behavior. The great advantage of the conditioning experiment is that it provides a method for defining learning strictly in stimulus-response terms. Eager to analyze even the most complex behavior into simple units, some psychologists have been willing to consider the conditioned response as typical of all learning. In this way, they were able to bridge the gap between animal and human behavior. Some of the most important systematic theories of learning today have been based upon conditioning concepts. However, as experimentation proceeded in the laboratory, conditioning turned out to be a much more complex process than the original investigators had believed. It is now clear that the establishment of a conditioned response, both in lower animals and in man, is a function of a large number of variables. Of course, some of these variables are crucial to conditioning; others merely serve to modify the course of the conditioning process.

A Typical Conditioning Experiment

Let us now look at a typical conditioning experiment as it was performed innumerable times in the laboratory of the Russian physiologist, I. P. Pavlov, the father of conditioning.

We find a dog, strapped comfortably in a harness, standing on the experimental table. The dog is thoroughly familiar with the experimental situation and does not object to the confinement. The room is soundproofed against external noises and the experimenter observes the animal, through a glass partition, from an adjoining room so as not to disturb his subject. Hanging from the dog's cheek is a rubber tube which has been connected to the duct of a salivary

gland. A metronome sounds, the dog pricks up his ears and turns his head toward the sound. No saliva drops from the rubber tube. Several minutes later, a plate of meat powder is moved within reach of the dog's mouth. As he eats the food, several drops of saliva fall from the rubber tube and are collected in a sensitive measuring device. Some time after the meat has been removed, the metronome sounds again, and this time it is followed within 10 seconds by the plate of meat. At the sight of the food, the dog salivates. This pairing of the metronome and the meat is repeated over and over. On the fifth presentation of this combination of stimuli, two drops of saliva fall from the rubber tube after the sounding of the metronome but before the meat is presented. On the tenth sounding of the metronome, the meat powder is omitted entirely. The metronome sounds for 30 seconds and during this period eight drops of saliva are collected from the rubber tube.

The Main Concepts of Conditioning

We shall now examine the main concepts which have been used in the description and interpretation of such experiments.

The Unconditioned Stimulus and the Unconditioned Response. In order to be conditioned, the animal must initially have a uniform and consistent response to a particular stimulus. Thus, a dog uniformly and consistently salivates when there is food in his mouth. The stimulus which at the beginning of the experiment can be relied upon to evoke a particular response is called the *unconditioned stimulus* (US). The response elicited by the unconditioned stimulus is the *unconditioned response* (UR). This response is part of the innate equipment of the animial.

The Conditioned Stimulus and the Conditioned Response. As we have seen, Pavlov's experimental animal learned to salivate in response to a metronome. Here lies the crux of the conditioning process. A stimulus comes to elicit a response which it originally did not evoke. At first, the dog did not salivate at the sound of the metronome. Having been paired repeatedly with the meat powder, the metronome finally became a reliable stimulus for salivation. The sound of the metronome is the *conditioned stimulus* (CS), and the secretion of saliva is the *conditioned response* (CR). Any stimulus which comes to evoke the unconditioned response by virtue of being

paired with the unconditioned stimulus is called a conditioned stimulus. When the unconditioned response is made to the conditioned stimulus, it is called the conditioned response.

CR and UR Are Not Necessarily Identical. An important qualification has to be added to our description of the conditioning process. In many experimental situations, the conditioned response is different from the unconditioned response, both in its nature and intensity. For example, one of the unconditioned responses of the guinea pig to electric shock (US) is a sharp intake of the breath (UR). When this response was conditioned to an auditory signal (CS), a conditioned response was restrained breathing (CR). As we shall see later, the unconditioned response and the conditioned response may even involve different effector organs.

Experimental Extinction and Spontaneous Recovery. A conditioned response is established when the conditioned stimulus is associated with an unconditioned stimulus. When the unconditioned stimulus is later omitted, the conditioned response gradually disappears. Let us return for a moment to Pavlov's experiment. After the dog had learned to salivate at the sound of the metronome, the experimenter presented the metronome repeatedly but he failed to give food to the dog. As this procedure was continued, the salivation first decreased in volume and finally stopped altogether. The response had been *experimentally extinguished*. Usually the dog was then returned to his living quarters. On the following day, the experiment was resumed. When the metronome sounded again, the salivary response reappeared. The phenomenon is known as *spontaneous recovery*. Clearly, experimental extinction is different from forgetting. When a response is forgotten, it does not recur without relearning. When a response is extinguished, it will reappear under certain experimental conditions.

Generalization and Discrimination. When conditioning has taken place, the particular stimulus used as the conditioned stimulus is not the only stimulus which will elicit the conditioned response. Other stimuli, which were neutral before conditioning took place, also come to elicit the response. If the conditioned stimulus is a light of given intensity (say, i_5), then other lights of lesser intensity (i_1, i_2, i_3, i_4) and of greater intensity (i_6, i_7, i_8, i_9) will also elicit the response, though less frequently than the original conditioned stim-

ulus, i_5. The fact that conditioned responses are made to stimuli other than the specific conditioned stimulus is known as *generalization*. The greater the difference between the conditioned stimulus and the stimulus used to test for generalization, the less frequently will the response be elicited. Thus, i_1 and i_9 would produce conditioned responses less frequently than i_4 and i_6. In short, there is a *gradient* of generalization: the closer a test stimulus is to the original conditioned stimulus, the more effective it is in evoking the conditioned response. Generalization is most marked when the conditioned stimulus differs from the stimuli used to test for generalization along a single physical dimension, such as the intensity of light. (See also pp. 431 f.)

The degree of generalization can be reduced by presenting the unconditioned stimulus (e.g., food or shock) after responses to the conditioned stimulus proper, e.g., only after responses to i_5 but not after responses to other stimulus intensities. By this procedure, the animal is led to *discriminate* between the conditioned stimulus and other stimuli. The degree to which discrimination can be pushed provides us with a good measure of the animal's sensitivity to stimulus differences. The conditioning procedure has, in fact, been used as an equivalent to psychophysical methods to measure the differential thresholds of animals.

Sometimes the experimenter may make the task of discrimination too difficult for the animal, requiring it to make discriminations which exceed its sensory capacities, making, for example, the intensity differences between the conditioned stimulus and the test stimuli so small as to be below the animal's differential threshold. In such cases, the animal's behavior may be severely disrupted: it may show great agitation, try to break out of its harness, growl and snap at the experimenter, and become permanently incapable of serving as a subject in a conditioning experiment. Such disruption of behavior is aptly described as *experimental neurosis*. Experimental neurosis is induced when the animal in a conditioning experiment faces an impossible task, when the attainment of rewards and punishments depends on a fine discrimination which it is incapable of making. In anthropomorphic terms, it is put in a conflict between responding and not responding. Like many humans, who cannot resolve a conflict, it becomes "neurotic."

The Main Parameters of Conditioning Experiments

We have said that the establishment of a conditioned response and the form which the response takes depend on many variables. Nevertheless, we may distinguish between major determinants of the process and those factors which merely modify it. By major determinants, we shall mean those experimental procedures without which no conditioned response can be elicited. The modifying factors are those which affect the rate and degree of conditioning. We shall discuss the major determinants first.

Strength of Unconditioned Response. It is not sufficient for the establishment of a conditioned response that two stimuli occur together. As we have emphasized, the unconditioned stimulus must uniformly and consistently elicit the unconditioned response. It is for this reason that conditioned responses can always be traced back to innate response patterns—the most uniform and consistent responses in the repertory of the organism.

Nature of CS. The conditioning experiment tries to establish a connection between the CS and the UR. At the beginning of the experiment, therefore, the CS must be "neutral" with respect to the UR. This does not mean that initially the CS never evokes the UR at all. It does mean that at the start the CS should not evoke the UR uniformly and consistently. Nor should the original response to the CS be of such a nature as to compete seriously with the UR. Suppose, for example, the US is a pinprick to the *right forelimb* of a dog. The CS is a strong electric shock to the *right hindlimb*. The UR to the pinprick is a flexion of the right foreleg, while the original response to the electric shock includes flexion of the right hindleg. Under these circumstances, the dog may not learn to flex its forelimb in response to the electric shock. Apart from these restrictions, virtually any stimulus may serve as the CS.

Temporal Relations Between CS and US. The temporal interval between the conditioned stimulus and the unconditioned stimulus may vary. This interval, however, must not be too great or conditioning will not take place. Conditioning is most successful if the CS precedes the US by about 0.5 seconds. As the interval is changed from this optimal length, conditioning becomes more and more difficult. For a given temporal interval, it is easier to establish a condi-

tioned response when the CS *precedes* the US than when it follows it. Fig. 74 graphically shows how degree of conditioning varies with the temporal interval between CS and US.

It should be noted that conditioning apparently occurs even when the conditioned stimulus follows in time the unconditioned stimulus. The acquisition of a response under these particular temporal con-

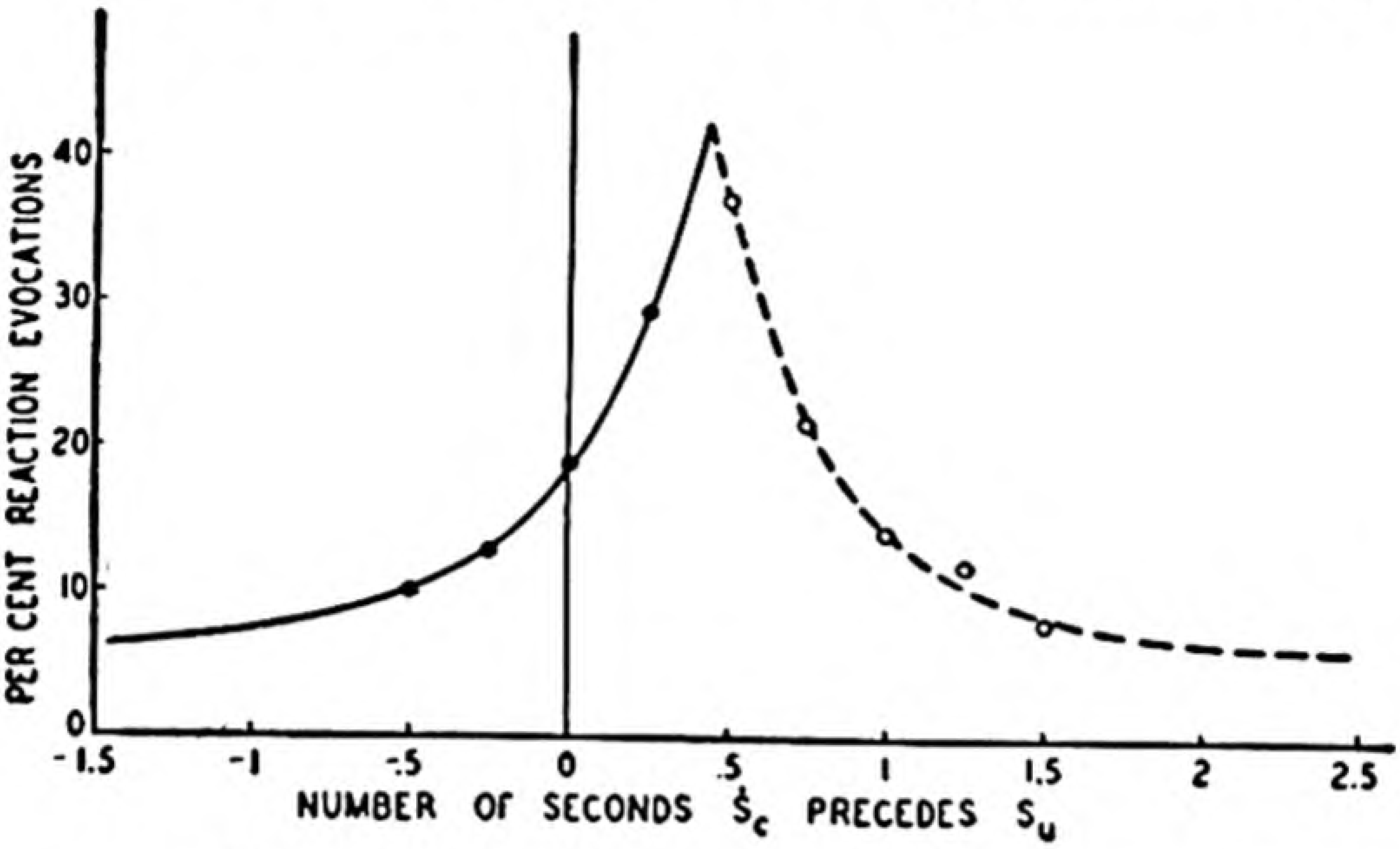


FIG. 74. Degree of Conditioning (percent of conditioned responses obtained) as a Function of the Time Interval Between the Conditioned Stimulus and the Unconditioned Stimulus. Negative time intervals indicate that the conditioned stimulus followed the unconditioned stimulus. (From C. L. Hull, *Principles of behavior,* 1943, p. 170, by permission of Appleton-Century-Crofts, Inc. From data by H. M. Wofle, Conditioning as a function of the interval between the conditioned and the original stimulus, *J. Gen. Psychol.,* 1932, 7:80-103.)

ditions is difficult to obtain, and some investigators have maintained that *backward conditioning* does not occur.

The CS may or may not overlap with the US in time. For example, the metronome may still be sounding when the food is presented in a Pavlovian experiment. In this case, we speak of a *delayed CR.* On the other hand, if the metronome is turned off some time before the food appears, we speak of a *trace CR.* Trace CR's are more difficult to form than delayed CR's. In both cases, a long time interval becomes itself a conditioned stimulus, and the CR tends to appear toward the end of the interval.

Number of Acquisition Trials. Like all learning, conditioning takes time. As we have seen, the CS and US must be presented together *repeatedly* before the CS will elicit the CR. Just as there are

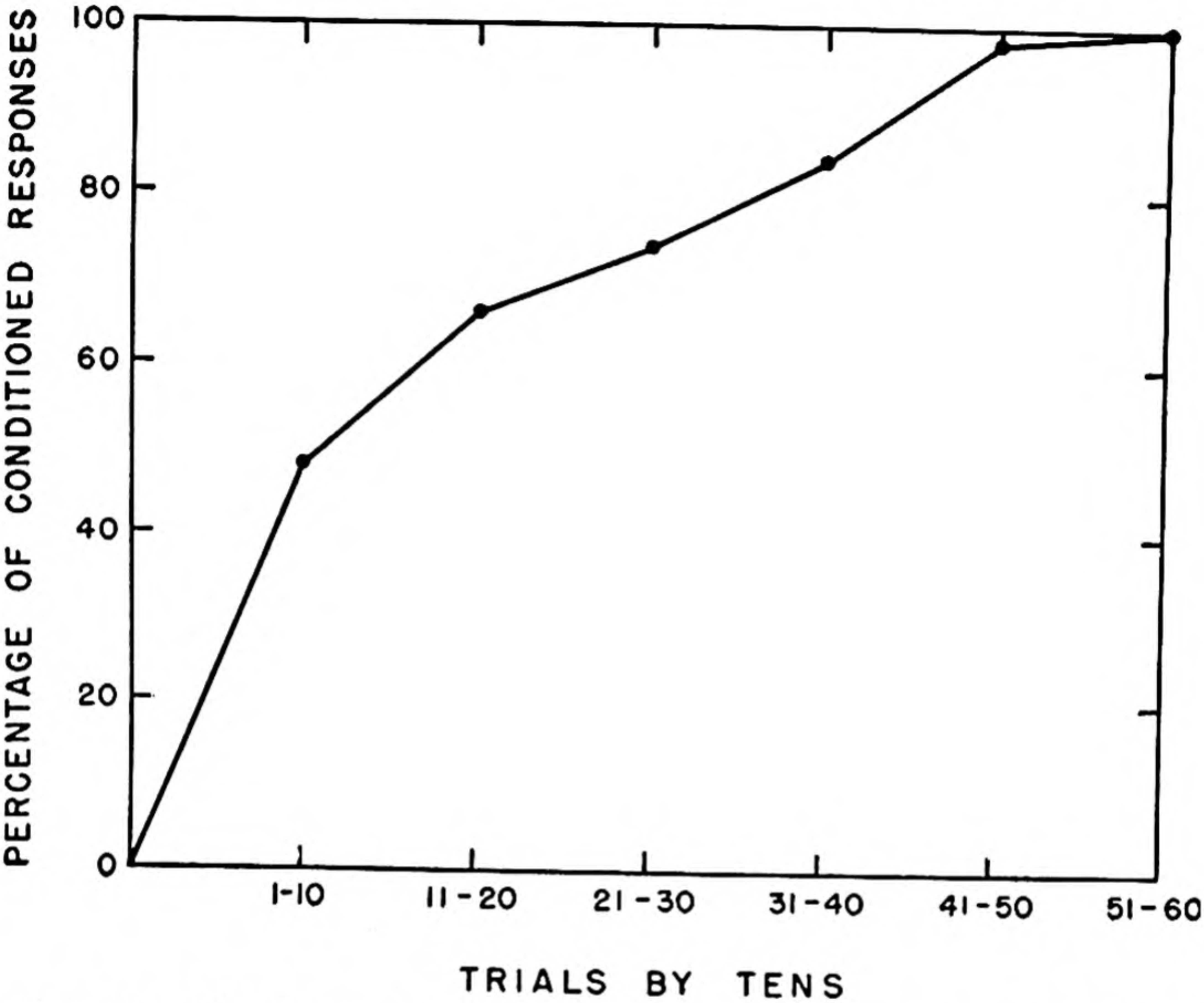


FIG. 75. This Curve Shows How a Conditioned Response (conditioned eyeblink of human subjects) Grows in Strength as a Function of the Number of Repetitions. (After G. A. Kimble, Conditioning as a function of the time between conditioned and unconditioned stimuli, *J. Exper. Psychol.*, 1947, *37*:8, by permission of the journal and the American Psychological Association.)

acquisition curves for verbal materials, so there are acquisition curves for conditioned responses. Fig. 75 shows how a conditioned response grows in strength as the number of repetitions increases.

Summary of Major Determinants. The establishment of a CR requires: (1) the presence in the subject of a reliable UR to a US; (2) initial neutrality of the CS with respect to the UR; (3) a close sequence in time of the CS and US; and (4) several paired presentations of CS and US. The experimental conditions must conform to

each of these factors. However, these factors are interdependent. The number of acquisition trials required for conditioning, for example, varies with the strength of the US. Fig. 76 shows that the

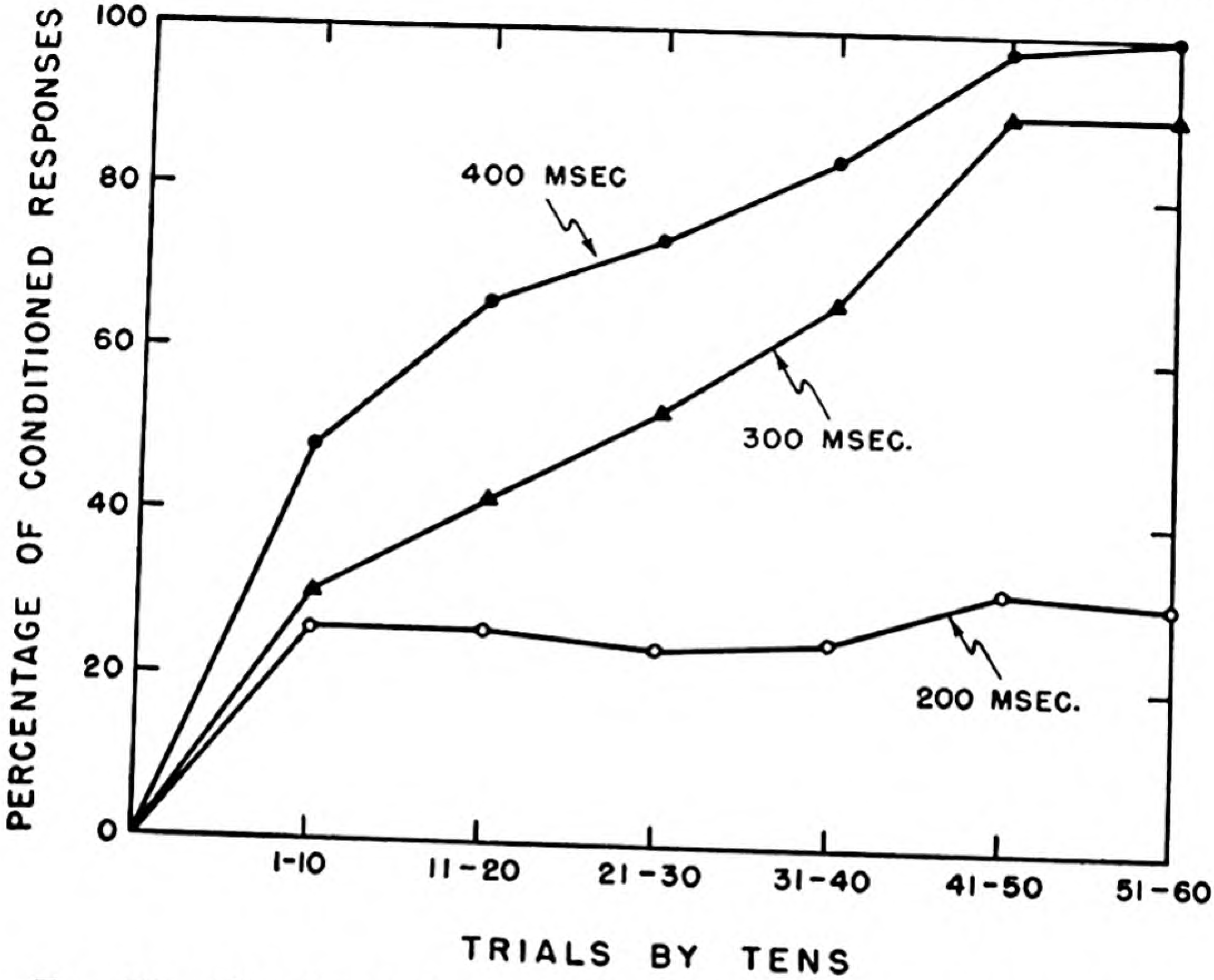


FIG. 76. The Dependence of the Speed of Conditioning on the Temporal Interval Between Conditioned Stimulus and Unconditioned Stimulus. These curves show the acquisition of a conditioned eyelid response with three temporal intervals between the conditioned stimulus and the unconditioned stimulus. (After G. A. Kimble, Conditioning as a function of the time between conditioned and unconditioned stimulus, *J. Exper. Psychol.,* 1947, 37:8, by permission of the journal and the American Psychological Association.)

number of acquisition trials also varies with the temporal interval between CS and US. In general, within strict limits, if one of the major determinants is relatively unfavorable to conditioning, this situation may be offset by more favorable values of the other parameters.

Secondary Determinants

Distribution of Practice. The number of acquisition trials per unit of time influences the degree of learning. In general, the more distributed the trials are in time, the better the learning. Fig. 77 shows that a conditioned eyelid response in man is acquired much

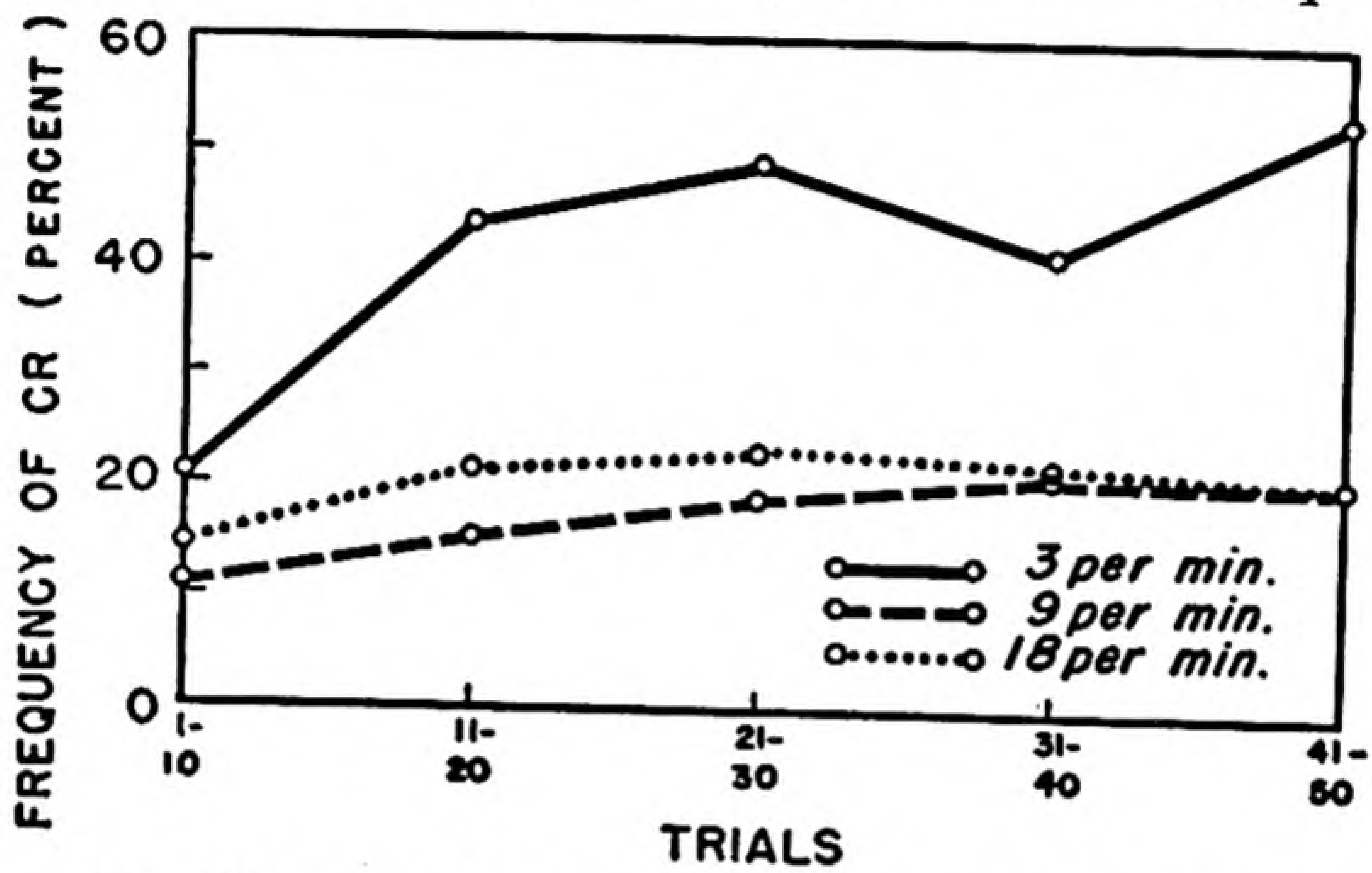


Fig. 77. The Influence of Distribution of Practice on the Acquisition of a Conditioned Eyelid Response. The conditioned stimulus and unconditioned stimulus were paired 50 times at the three rates indicated. The superiority of a slow rate is apparent. (From E. R. Hilgard and D. G. Marquis, *Conditioning and learning*, 1940, p. 149, by permission of Appleton-Century-Crofts, Inc.)

more readily when the stimuli are presented once every 20 seconds than if they follow each other at a faster rate.

Percentage of Trials in Which US Appears. For the establishment of a CR, the US must be paired with the CS. However, the US need not accompany the CS on every trial. A reliable CR can be established even if the US appears on only 50 percent of the trials. Only further experiments will discover for what fraction of the trials the US needs to appear.

External Inhibition. In the experiment which we described, the animal was kept in a soundproofed, screened room for a good reason. An extraneous stimulus that acts while the animal is acquiring the CR may temporarily decrease the strength of the response.

To this phenomenon, the term *external inhibition* has been applied. The effect of the extraneous stimulus, characteristically, is of short duration. An extraneous stimulus may also have the opposite effect (*disinhibition*). If this stimulus occurs immediately after the CR has been experimentally extinguished, the response may temporarily reappear.

Individual Differences. Conditioned responses, like all the activities of living organisms, show individual differences. Animal and human subjects alike differ in the ease with which they can be conditioned. Little is known as yet about the constitutional and attitudinal factors responsible for these differences.

Attitudinal Factors. Whenever the UR is under voluntary control (human subjects), even if only partially, attitudinal factors may assume a major role. Much will depend on how the subject interprets the task set him by the experimenter. If a shock to the hands is the US, for example, the subject may consider it a test of endurance, or a test of the speed with which he can achieve the withdrawal. Attitudinal factors need to be controlled carefully when human subjects are involved.

Types of Conditioning Experiments

The term *conditioning* was originally restricted to the procedures which we have just described. The connotation of the term was, however, soon extended to encompass a greater variety of learning situations. The essential differences among the various conditioning experiments reside in the role played by the subject's motivation. In the Pavlovian experiment, an adequate degree of motivation is insured by the use of an unconditioned stimulus which elicits a strong innate response. The unconditioned stimulus is paired with the conditioned stimulus no matter what the animal does. However, conditioned responses can also be established in situations where the animal has to discover by a process of trial and error the means of bringing about the unconditioned stimulus. In these trial-and-error situations, motivating factors play a crucial part in that they lead to the *selection of the response* that finally becomes conditioned. It thus becomes necessary to distinguish between two general types of conditioning procedures.

Classical Conditioning. When the experimental conditions are

Food Bell

such that the US accompanies the CS no matter what the subject does, a *classical conditioning* experiment is performed. This procedure is called classical conditioning because it represents the procedure used by Pavlov. Since the CR markedly resembles the response (UR) evoked by the US, the application of the US is termed *homogeneous reinforcement*. The term *reinforcement* is used because the presentation of the US strengthens, or reinforces, the CR. The adjective *homogeneous* refers to the fact that CR and UR are similar activities, involving the same effector organs, e.g., salivation.

Instrumental Conditioning. When the experimental conditions are such that the application of the reinforcing stimulus depends upon what the animal does in response to the CS, an *instrumental conditioning* experiment is performed. It is so called because the response to the conditioned stimulus is instrumental in producing the reinforcing stimulus. In this case, the conditioned response does not resemble the response evoked by the unconditioned stimulus. For this reason, we speak of *heterogeneous reinforcement*.

Conditioning an Instrumental Response

In order to illustrate the distinction between classical and instrumental conditioning, we shall describe a situation in which a subject learns an instrumental response. Again, we find a dog, strapped in his harness, standing on the experimental table but minus the tube dangling from his cheek. A bell sounds, the animal fails to respond to it but continues to lick his paw. No food is given to him. The bell sounds again, and now the dog pricks up his ears and turns toward the bell. This time food is immediately presented, and he nibbles from it. The bell sounds again and again, and only when the dog pricks up his ears and turns toward the sound, is he rewarded with food. After several trials, no sooner has the bell begun to sound than the dog's ears go up promptly, he turns his head with mechanical precision toward the source of sound, and saliva drops from his mouth. An instrumental response—pricking up the ears and turning the head—has been learned. But note: we have at the same time unwittingly performed a classical conditioning experiment. We merely failed to measure the increase in salivation as the experiment proceeded. The instrumental response and the

classical response have both been established through the same reinforcing stimulus—food. One response—salivation—has been established by homogeneous reinforcement, the other—the instrumental one—by heterogeneous reinforcement.

Four main types of instrumental conditioning have been distinguished according to the consequences of the instrumental response for the motivated subject.

Reward Training. The instrumental response leads to a reward. For example, a buzzer is sounded and food is delivered to the animal only if it performs some instrumental act such as pressing a lever or turning its head toward the sound. The appearance of the reward (reinforcing stimulus) is contingent upon the instrumental act.[1]

Escape Training. The instrumental response allows the subject to escape a harmful or painful stimulus. For example, the animal may terminate a continuous electric shock by an instrumental act such as pressing a pedal or standing upright. The well-conditioned animal will perform this instrumental act very shortly after the onset of the noxious stimulus.

Avoidance Training. The instrumental response enables the subject to avoid the onset of a harmful or painful stimulus. For example, a buzzer is sounded and followed after a set time interval by an electric shock. By performing an instrumental act such as withdrawing from the electrodes during this time interval, the subject may avoid the application of the shock.

Secondary Reward Training. The instrumental response leads to a stimulus object which may later be exchanged for a reward. In other words, the subject works for a token. This token is originally neutral and acquires reinforcing properties because it is paired with a reward. For example, an animal will learn to pull a string with a weight attached to it in order to obtain a poker chip or similar token, provided the animal can use the token to get food.

[1] Sometimes no specific conditioned stimulus, such as a buzzer, is given, and the instrumental response (e.g., pressing a lever) must occur spontaneously before reinforcement is given. In such cases, we may think of the instrumental response as conditioned to the total pattern of stimuli in the experimental situation.

Quantitative Methods in Conditioning

Let us reconsider the main independent and dependent variables of the conditioning experiment with special emphasis on the methods of quantification to which they lend themselves.

Independent Variables

The principal independent variables in a conditioning experiment are: (1) the nature and intensity of the US in relation to the motivation and the physical state of the subject; (2) the nature and intensity of the CS; (3) the time interval between the CS and the US or between the CR and the US; (4) the interval between successive trials; (5) the number of acquisition trials; (6) the relative frequency of reinforcements.

The Nature and Intensity of the US. We have emphasized that conditioning depends on the subject's uniform, consistent response to the US. The degree or strength of the conditioned response depends on the intensity of the US because this intensity in turn determines the strength of the UR. It is probably safe to say that within wide limits, the strength of conditioning is highly correlated with the intensity of the US. In many cases, however, if the intensity of the US is increased beyond a certain point, the behavior sequence may be disrupted.

It is not sufficient to know the strength of the US in order to predict the strength of the UR. The efficacy of the US depends on the motivation and physical state of the subject. When an animal is satiated, a large food reward is no more effective as a reinforcing stimulus than a small reward. A constant reward will have different effects depending on the motivation of the animal.

Nature and Intensity of the CS. As long as the CS lies within the sensory capacity of the organism and is neutral with respect to the UR, it raises no special problems of quantification. In describing a CR experiment, it is usually sufficient to give an adequate description of the CS in physical terms.

Time Interval Between CS and US. Although this variable is an important determinant of the strength of conditioning, there is again no special problem of quantification because it is stated in

units of physical time. In instrumental conditioning, the time interval between CR and US may become a critical variable.

Time Interval Between Successive Trials. The same considerations apply as to the question of the interval between CS and US although laboratory control is usually achieved without difficulty.

Number of Acquisition Trials. The number of acquisition trials is a simple counting variable. This means that we are simply concerned with the frequency of occurrence rather than with the intensity or form of a given event.

Relative Frequency of Reinforcement. This variable refers to the percentage of the total number of trials on which the reinforcing stimulus is presented. Again, we are dealing with a simple counting variable. Strength of conditioning can be plotted as a function of the relative frequency of reinforcement.

Dependent Variables

Two features of the conditioned response constitute the main dependent variables: the *form* of the CR and the *strength* of the CR. One of these, the form of the CR, cannot be quantified. If in the course of conditioning, a response becomes more and more precise and circumscribed as is typically the case in withdrawal from shock, it is difficult to find a meaningful quantitative index of this change in the form of the response.

On the other hand, the strength of a CR *can* be rigorously quantified. Four important ways in which the strength of the CR can be measured are: (1) the frequency of occurrence of CR; (2) resistance to experimental extinction, i.e., the number of unreinforced trials required to reach a given criterion of experimental extinction; (3) the latency of the CR; and (4) the amplitude of the CR.

Frequency of Occurrence of CR. This variable is perhaps the most important single measure of the strength of the CR. It is again a simple counting variable. The experimenter must decide in advance on unequivocal requirements which a response has to meet in order to be considered a CR. Once this decision has been made, it is merely necessary to determine the number of CR's that occur in a specified block of trials.

Resistance to Experimental Extinction. Once a CR has been established, it can be extinguished if the reinforcement is omitted.

When we define strength of CR in terms of resistance to extinction, we mean that a strong CR requires more trials for extinction than a weak one. This index has been applied very widely in the measurement of instrumental reward conditioning.

Latency of the CR. Frequency of occurrence and resistance to extinction do not tell the whole story. The latency of the CR provides one supplementary measure. Under certain experimental conditions, a short latency signifies a strong CR. One important exception to this rule shows why latency is not a universal measure. In the case of the delayed CR and the trace CR, latency actually increases as the response becomes more strongly established. The well-established CR of this type occurs in close temporal proximity to the US.

Amplitude of the CR. Again, under certain experimental conditions only, the amplitude of a response can be used to gauge its strength. We then say that the greater the strength of the CR, the greater the amplitude of the response, e.g., excursion of the limb. Again we may cite an important exception to show that this rule is not universal. An animal's first response to a strong electric shock is violent and diffuse. It may be described as a response of large amplitude. As the animal learns to avoid the shock, the response becomes more localized and precise. When learning is complete and the response has reached its maximally adaptive stage, the CR has only a fraction of its original amplitude.

Special Problems of Control in Conditioning Experiments

The basic notion of conditioning is simple, but many delicate problems of control arise in any conditioning experiment. Without attempting to be exhaustive, we shall now discuss some of these special problems.

Control of External Environment. In many psychological experiments, only certain relevant environmental features need to be carefully controlled. For example, one can determine the laws of color mixture in a room which is not absolutely quiet. In the conditioning experiment, on the other hand, even minute disturbances can interfere with the learning process. We have already referred to Pavlov's concept of external inhibition in connection with such

effects. Furthermore, some modern writers have suggested that *any* stimulus acting at the same time as the US will become associated with the UR. After all, the CS does not come with a tag on it. On any given trial, the animal cannot know what stimulus is extraneous and what stimulus is intended as the CS. For this reason, experimenters take great care to isolate the subject as much as possible during

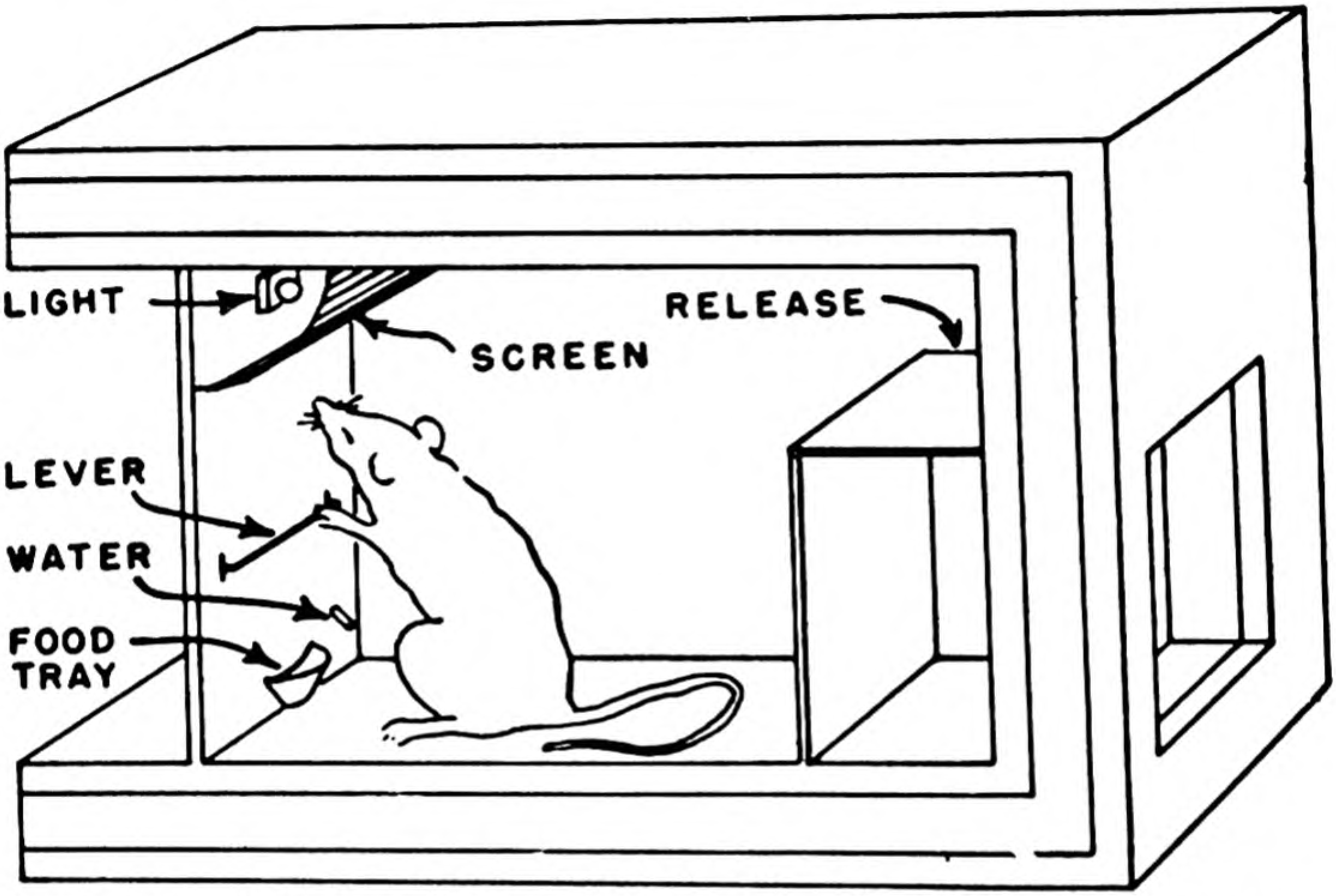


FIG. 78. "Skinner Box." This is widely used in instrumental conditioning experiments. The rat is well isolated in this soundproofed chamber. When the rat presses the lever, a pellet is automatically delivered from a food magazine into the tray. (After B. F. Skinner, *The behavior or organisms: an experimental analysis,* 1938, p. 49, by permission of Appleton-Century-Crofts, Inc.)

a conditioning experiment. One device which has been used with great success is the Skinner box. The subject (e.g., a white rat) is isolated in a small, dark, soundproof chamber, and in this way the amount of disturbance during the conditioning trials is greatly reduced. Fig. 78 shows such a box. This device is being widely used in the quantitative investigation of instrumental conditioning.

Control of Subject's Motivation. The effectiveness of the reinforcing stimulus is not determined by its physical intensity alone, but varies significantly with the subject's motivation. Two sources of reinforcement, which have been widely used in conditioning, will serve to illustrate this point: food and shock. Food is an effective reinforcing stimulus only if the animal is sufficiently hungry. The

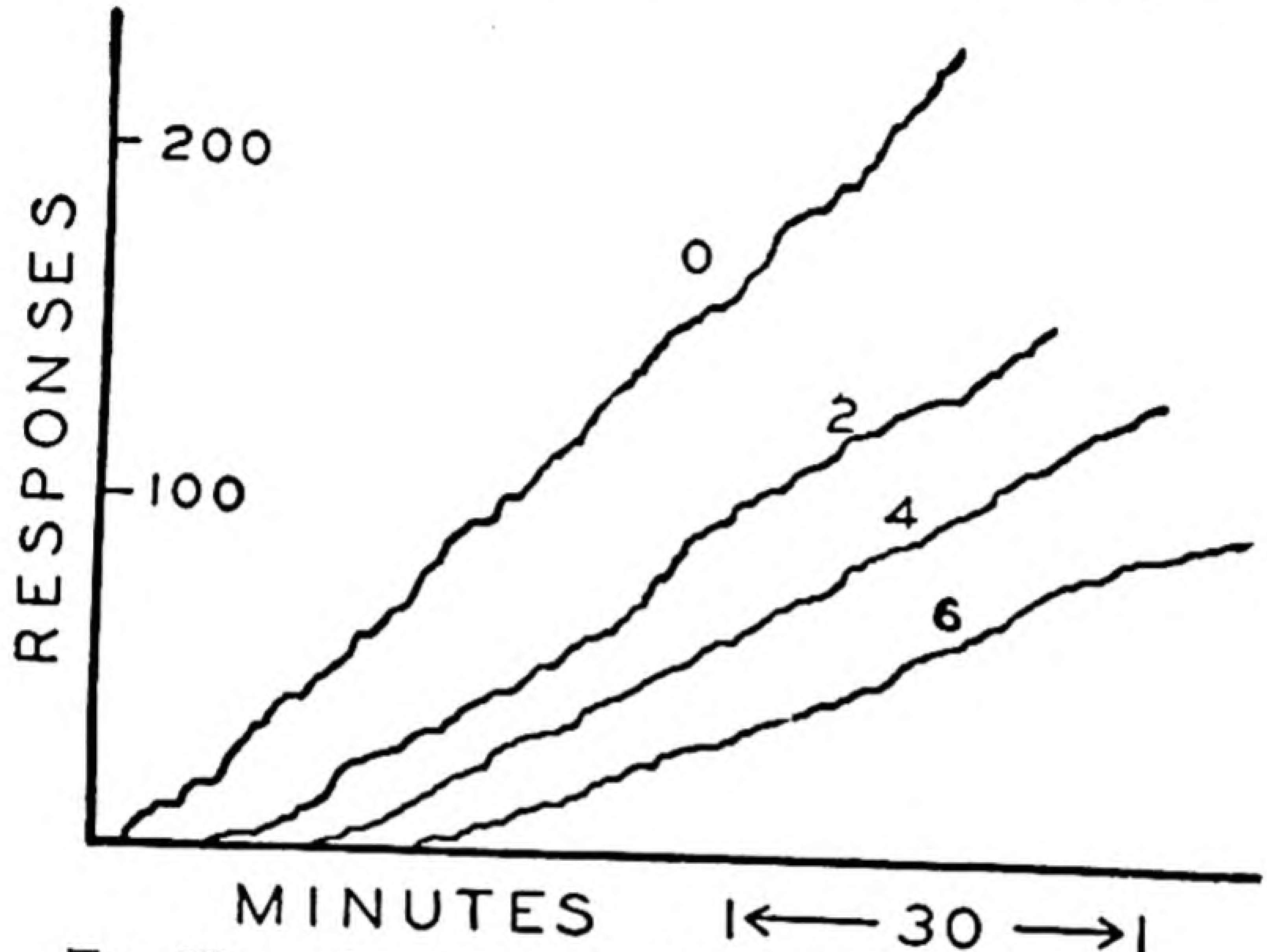


FIG. 79. These Curves Show How a Rat's Rate of Lever-Pressing Response Is Influenced by Degree of Hunger. The numbers on the curves refer to grams of food eaten before the experimental session. The more food has been eaten, the fewer the number of lever-pressings per unit of time. (From B. F. Skinner, *The behavior of organisms: an experimental analysis,* 1938, p. 393, by permission of Appleton-Century-Crofts, Inc.)

degree of the subject's hunger can be manipulated by the simple procedure of feeding and fasting. Over a certain range, the greater the degree of hunger, the more quickly the conditioned response is established. Once the response has been established, the degree of hunger will help determine the readiness with which it is used. The frequency with which an animal responds in an instrumental situation may vary with the degree of hunger. Fig. 79 illustrates this dependence of rate of response on hunger.

A noxious stimulus, such as electric shock, behaves in a similar

way: the effectiveness of shock in the conditioning process varies with the subject's sensitivity to it. Holding the physical intensity of the shock constant from trial to trial does not insure constant reinforcement. Under some conditions, it is necessary to increase the intensity of the shock in order to maintain the response at a constant level; under other conditions, the shock must be decreased in intensity. Thus, two factors which lead to changes in the effectiveness of shock are habituation to the punishment and changes in skin resistance (perspiration). These two factors may work in opposite

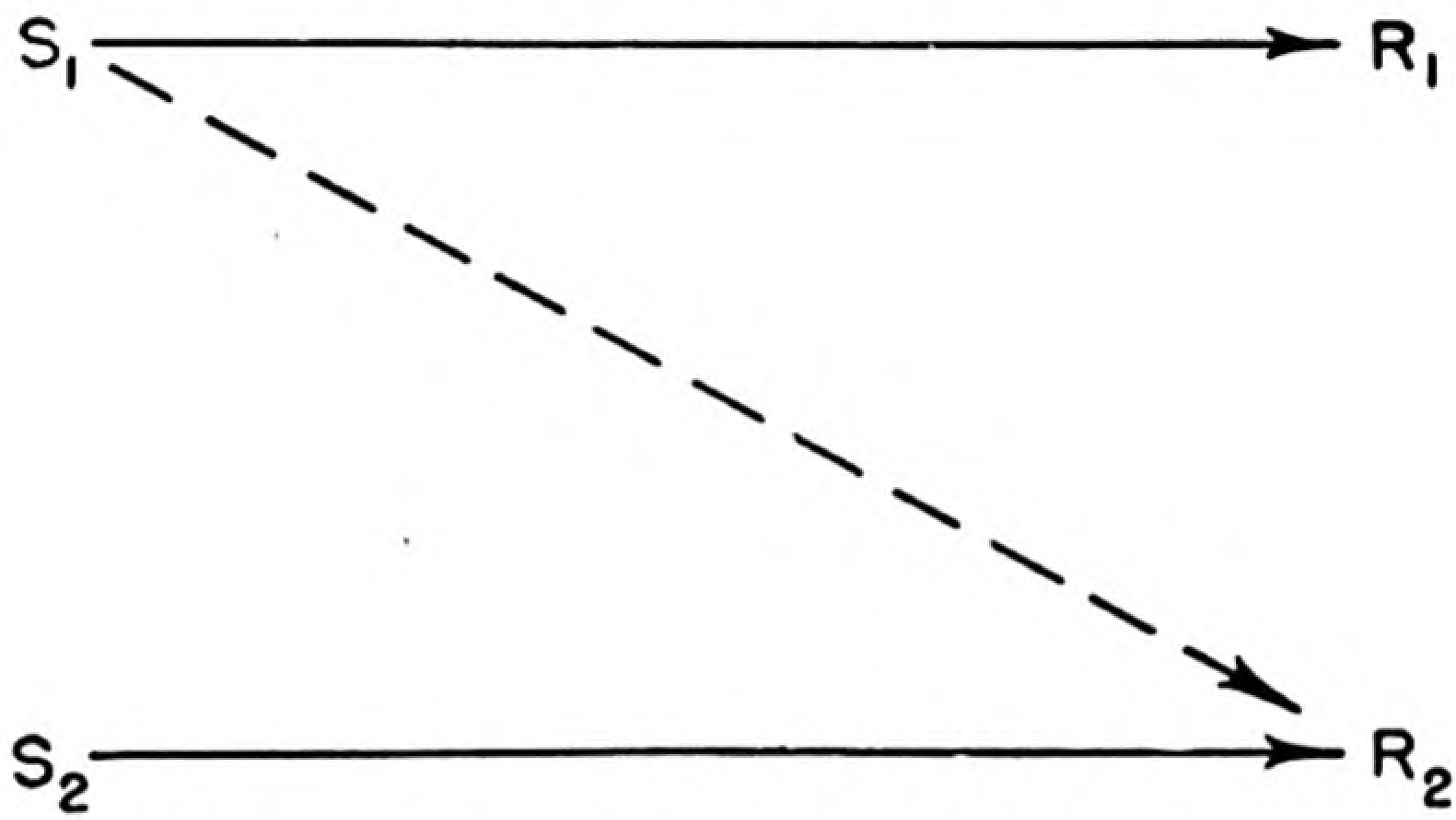


FIG. 80. A Simplified Diagram of the Conditioned Response. In the course of conditioning S_1 (CS) comes to evoke R_2 (UR).

directions. Furthermore, individual differences among subjects are so great that they need to be taken carefully into account.

Control of the Time Interval Between CS and US. The importance of this factor has already been emphasized. It leads to a special problem in one form of instrumental training: *avoidance conditioning*. In this training procedure, the situation is so arranged that the CR enables the subject to avoid punishment. Thus, a subject may avoid the shock by withdrawing his hand a short time after presentation of a CS. The time interval between CS and US must be sufficiently long to allow for the latency of the CR. Clearly, it takes time for the CR to occur after the CS has been presented (as much or more time as is required for a voluntary reaction). If the US follows the CS too quickly, the subject cannot avoid shock, and the purpose of the training procedure is defeated.

These are but a few of the problems of control which arise in conditioning experiments. They serve to emphasize the delicate nature of the conditioning process and to point up the danger of portraying it in terms of a simple diagram as the one reproduced in Fig. 80. This diagram tells us, correctly enough, that temporal pairing of stimuli is essential. It fails, however, to indicate the complex way in which many variables determine the establishment of a CR.

EXPERIMENT XIX
CONDITIONED HAND WITHDRAWAL

The experiment which we shall now outline calls for the conditioning of human subjects. We have already emphasized the fact that attitudinal factors are of major importance in this type of situation. Whether or not the unconditioned response is partially under voluntary control, the subject's *set* to inhibit the response or to facilitate it affects the degree and persistence of conditioning. It is always well to make the instructions explicit. The experimenter should not assume that his subject will take a "neutral set" if he does not instruct him.

This experiment illustrates a special case of instrumental conditioning, viz., *avoidance conditioning*. The term avoidance refers to the fact that the conditioned response leads to the avoidance of a noxious stimulus (electric shock) rather than to the presentation of a reward.

Purpose. To condition the hand-withdrawal response of a human subject. The reinforcing stimulus is an electric shock, a stimulus which generally elicits withdrawal responses in human subjects. The subject is to learn to avoid the shock by withdrawing his hand at a signal which precedes the noxious stimulus. Hand withdrawal, therefore, is the response instrumental in the avoidance of pain.

Apparatus. The experiment requires: (1) a source of electric shock; (2) a relatively neutral stimulus, such as a buzzer, which serves as the CS; (3) a device for presentation of the stimuli in a fixed temporal sequence; and, if possible, (4) an instrument for recording the withdrawal movement.

1. *Source of electric shock.* Any device which will deliver a controllable electric shock to the subject can be used. *The maximum electric shock which this instrument can deliver should not be of injurious intensity.* A typical source of shock consists of dry cells, an inductorium, and a pair of electrodes. Fig. 81 shows a schematic diagram. An alternative source of shock may be obtained from a condensor discharge. The

electrodes consist of two plates of polished metal, e.g., brass. A convenient size for the larger electrode is 3 inches square; the smaller one may be about 1 inch square. The larger electrode is mounted so that the palm of the hand may rest comfortably on it. The other,

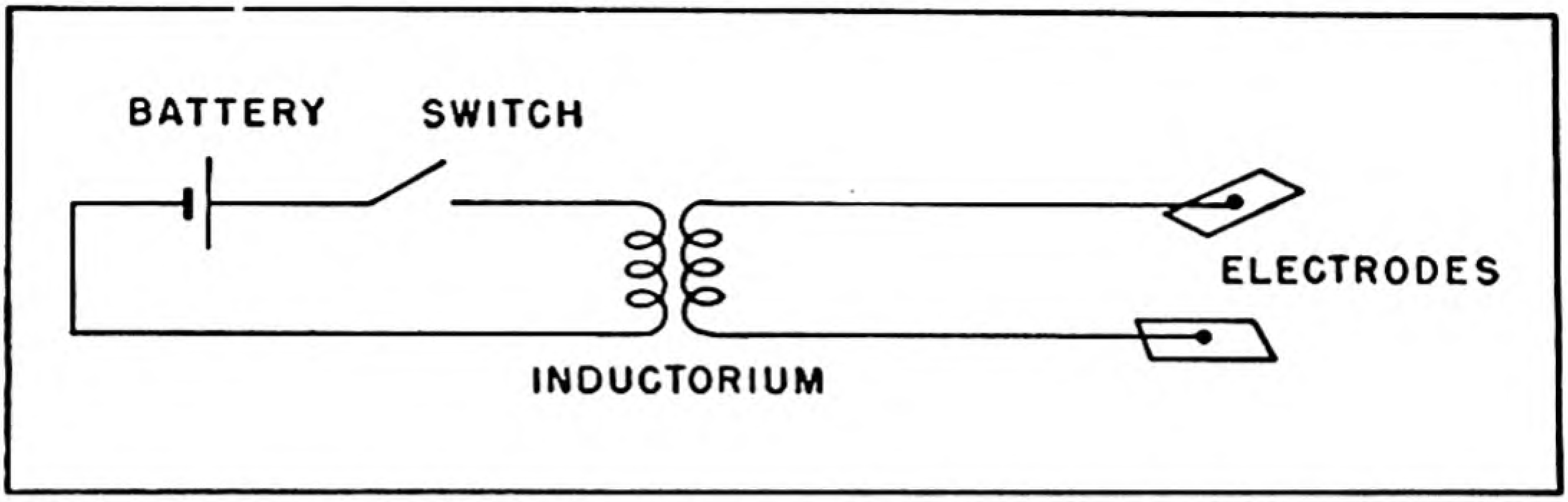


FIG. 81. Schematic Diagram of an Electrical Circuit for Administration of Electric Shock in Conditioned Hand-Withdrawal Experiments.

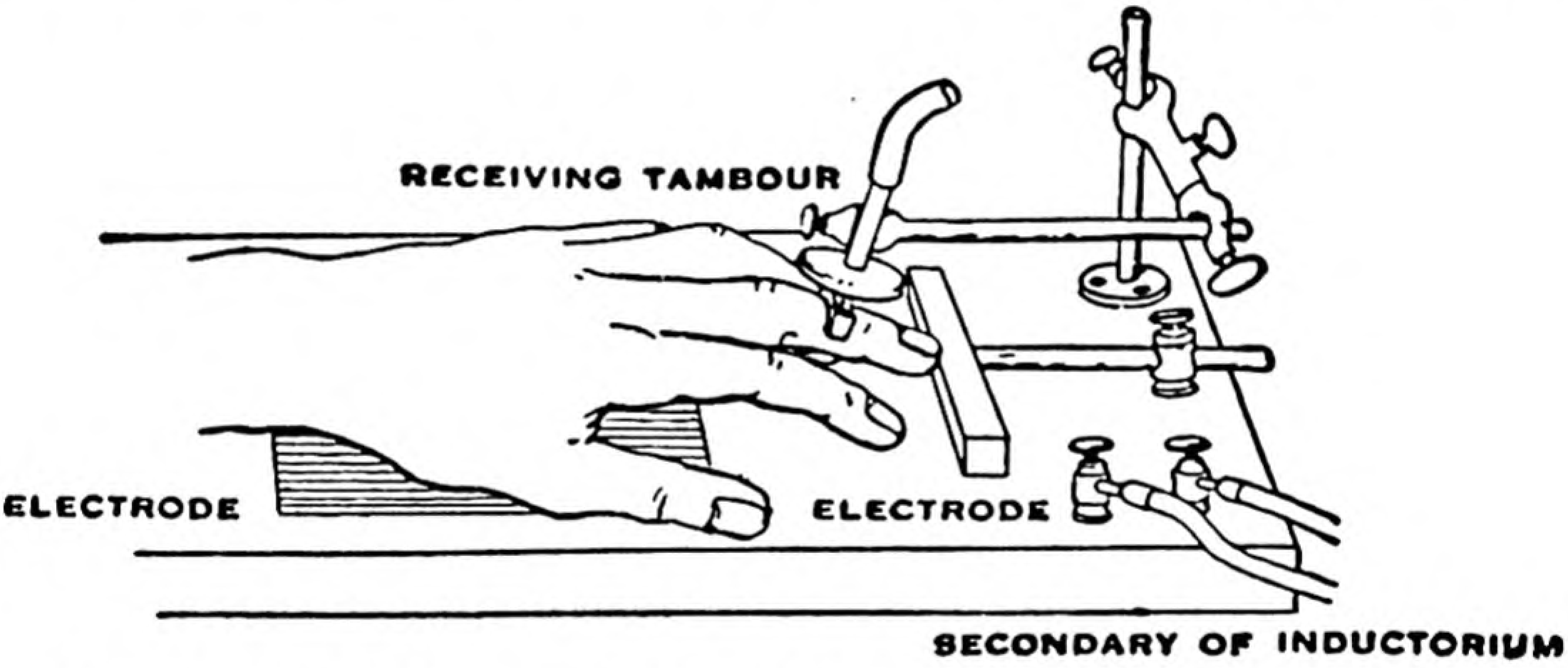


FIG. 82. This Figure Illustrates the Position of the Subject's Hand on the Electrodes in a Conditioned Hand-Withdrawal Experiment. (From J. B. Watson, *Psychology from the standpoint of the behaviorist,* 1919, p. 33, by permission of J. B. Lippincott Company.)

smaller, electrode is mounted so that the ball of the middle finger may be placed against it. Fig. 82 illustrates how the hand is placed on the electrodes. Care should be exercised that all electrical connections are firmly made.

2. *Conditioned stimulus.* Any stimulus which does not itself elicit a hand-withdrawal response may be used as the CS. A buzzer may be

employed but care should be taken that it is not startling to the subject. The click of a telegraph sounder, or a light, will serve equally well.

3. *Presentation of stimuli.* A device is needed to present the stimuli in fixed temporal sequence. For example, a pendulum which in the course of its fall successively closes two trip switches is adequate. A uniformly rotating disk may be used instead of a pendulum. The timing device should, if possible, deliver the CS about 0.5 seconds before the US. Of course, if the interval is made much shorter than 0.5 seconds, the subject will not be able to avoid the shock. If no mechanical device for presentation of the stimuli is available, the experimenter can operate two manual switches in quick succession and achieve adequate results.
4. *Recording device.* It is desirable to record on a polygraph or kymograph the occurrence of the stimuli and the responses. It is not necessary for the purpose of this experiment to make a record of the exact time at which these events occur. Only the presence or absence of hand withdrawal is critical. If such mechanical recording devices are not at hand, the procedure may be altered somewhat. Since the US is omitted on the test trials, the experimenter may simply note whether or not the withdrawal response occurs.

Procedure. The subject is seated comfortably at a table. The experimenter demonstrates to him how to place palm and middle finger on the two electrodes. The following instructions are then read:

"A series of electric shocks will be administered to you. We need at first to determine what intensity of shock you can tolerate without serious discomfort. Place your hand on the electrode as demonstrated, and I shall gradually increase the intensity of the shock. Tell me when the shock is strong enough to cause your hand to withdraw automatically. Relax as well as you can and do not resist the impulse to withdraw your hand."

In making this determination, it is important not to increase the intensity while a shock is being administered. Rather, a discrete series of shocks is presented, each one more intense than the preceding one. After the determination, the experimenter continues with the instructions:

"We shall now proceed to the experimental series. Some of the shocks will be preceded by a buzzer (light, or other CS). You are to withdraw your hand as soon as you feel the electric shock. Replace your hand on the metal plates immediately. This is not a test of your ability to endure shock. Be sure, therefore, that you withdraw your hand when you feel the electric shock. Sometimes, the electric shock

will not be preceded by the buzzer; at other times, the buzzer will not be followed by the electric shock. At no time should you try to force your hand to stay on the metal plates."

After having read the instructions, the experimenter presents the series of stimuli. At first, the experimenter presents the buzzer alone a few times (say, six) in order to establish the neutrality of the buzzer with respect to the withdrawal response. The interval between stimuli should be varied slightly from trial to trial and should be about 15 seconds. The experimenter then presents the shock alone for about the same number of trials. This is followed by a second series of the CS by itself. Although buzzer and shock have never been paired, some subjects may already respond to the buzzer by hand withdrawal. If they do, there has been *pseudo-conditioning* which is probably due to the sensitizing effect of the shock. If pseudo-conditioning has taken place, the CS should be presented alone until the subject does not respond for several trials in succession. When the experimenter is certain that the subject is not responding to CS alone, he begins to pair buzzer with shock. The interval between pairs should again be about 15 seconds. As the experimenter continues to pair buzzer and shock, he intersperses test trials at irregular intervals. A test trial consists of the presentation of the CS alone. Probably about every sixth trial, there should be a test trial. However, in order to prevent the subject from expecting a test, these trials should be irregularly spaced. If possible, the paired presentations are continued until the subject reaches a predetermined criterion, say CR's on five successive test trials.

After such a criterion has been reached, the experimenter tries to extinguish the response. For this purpose, he repeatedly presents the buzzer without shock. Again the procedure is continued until a predetermined criterion is reached, say, five successive failures to respond.

General References

Hilgard, E. R., The relationship between the conditioned response and conventional learning experiments, *Psychol. Bull.*, 1937, *34*:61-102.

Hilgard, E. R., *Theories of learning*, New York: Appleton-Century-Crofts, 1948.

Hilgard, E. R., and Marquis, D. G., *Conditioning and learning*, D. Appleton-Century Co., 1940.

Hull, C. L., Learning: II. The factor of the conditioned reflex. In C. Murchison (ed.), *A handbook of general experimental psychology*, Worcester: Clark University Press, 1934.

Hull, C. L., *Principles of behavior*. New York: D. Appleton-Century Co., 1943.

Pavlov, I. P., *Lectures on conditioned reflexes* (Trans. by W. H. Gantt), New York: International, 1928.

Skinner, B. F., *The behavior of organisms*, New York: D. Appleton-Century Co., 1938.

Spence, K. W., Theoretical interpretations of learning. In F. A. Moss (ed.), *Comparative psychology*, New York: Prentice-Hall, 1942.

Nature of Conditioned Stimulus and Response

Culler, E., Recent advances in some concepts of conditioning, *Psychol. Rev.*, 1938, *45*:134-153.

Hilgard, E. R., The nature of the conditioned response: I. The case for and against stimulus-substitution, *Psychol. Rev.*, 1936, *43*:366-385.

Hilgard, E. R., The nature of the conditioned response: II. Alternatives to stimulus-substitution, *Psychol. Rev.*, 1936, *43*:547-564.

Schlosberg, H., The relationship between success and the laws of conditioning, *Psychol. Rev.*, 1937, *44*:379-394.

Acquisition and Extinction

Brogden, W. J., and Culler, E., Experimental extinction of higher-order responses, *Amer. J. Psychol.*, 1935, *47*:663-669.

Brogden, W. J., Lipman, E. A., and Culler, E., The role of incentive in conditioning and extinction, *Amer. J. Psychol.*, 1938, *51*:109-117.

Finch, G., and Culler, E., Relation of forgetting to experimental extinction, *Amer. J. Psychol.*, 1935, *47*:656-662.

Hilgard, E. R., and Campbell, A. A., The course of acquisition and retention of conditioned eyelid responses in man, *J. Exper. Psychol.*, 1936, *19*:227-247.

Humphreys, L. G., The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions, *J. Exper. Psychol.*, 1939, *25*:141-158.

Hunter, W. S., Learning curves for conditioning and maze learning, *J. Exper. Psychol.*, 1936, *19*:121-128.

Kimble, G. A., Conditioning as a function of the time between conditioned and unconditioned stimuli, *J. Exper. Psychol.*, 1947, *37*:1-15.

Williams, S. B., Resistance to extinction as a function of the number of reinforcements, *J. Exper. Psychol.*, 1938, *23*:506-522.

Youtz, R. E. P., Reinforcement, extinction, and spontaneous recovery in a non-Pavlovian reaction, *J. Exper. Psychol.*, 1938, *22*:305-318.

Generalization and Discrimination

Hovland, C. I., The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone, *J. Gen. Psychol.*, 1937, *17*:125-148.

Hovland, C. I., The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone, *J. Genet. Psychol.*, 1937, *51*:279-291.

Hovland, C. I., The generalization of conditioned responses: III. Extinction, spontaneous recovery, and disinhibition of conditioned and of generalized responses, *J. Exper. Psychol.*, 1937, *21*:47-62.

Hovland, C. I., The generalization of conditioned responses: IV. The effects of varying amounts of reinforcement upon the degree of generalization of conditioned responses, *J. Exper. Psychol.*, 1937, *21*:261-276.

Lashley, K. S., An examination of the "continuity" theory as applied to discriminative learning, *J. Gen. Psychol.*, 1942, *26*:241-265.

Liddell, H. S., Conditioned reflex method and experimental neurosis. In J. McV. Hunt (ed.), *Personality and the behavior disorders*, Vol. II, New York: The Ronald Press Co., 1944.

Spence, K. W., The nature of discrimination learning in animals, *Psychol. Rev.*, 1936, *43*:427-449.

Spence, K. W., The differential response in animals to stimuli varying within a single dimension, *Psychol. Rev.*, 1937, *44*:430-444.

Temporal Relations

Bernstein, A. L., Temporal factors in the formation of conditioned eyelid reactions in human subjects, *J. Gen. Psychol.*, 1934, *10*:173-197.

Kappauf, W. E., and Schlosberg, H., Conditioned responses in the white rat: III. Conditioning as a function of the length of the period of delay, *J. Genet. Psychol.*, 1937, *50*:27-45.

Perin, C. T., A quantitative investigation of the delay-of-reinforcement gradient, *J. Exper. Psychol.*, 1943, *32*:37-51.

Wolfle, H. M., Time factors in conditioning finger-withdrawal, *J. Gen. Psychol.*, 1930, *4*:372-378.

Attitudinal Factors

Cole, L. E., A comparison of the factors of practice and knowledge of experimental procedure in conditioning the eyelid response of human subjects, *J. Gen. Psychol.*, 1939, *20*:349-373.

Grant, D. A., The influence of attitude on the conditioned eyelid response, *J. Exper. Psychol.*, 1939, *25*:333-346.

Hilgard, E. R., An algebraic analysis of conditioned discrimination in man, *Psychol. Rev.*, 1938, *45*:472-496.

Hilgard, E. R., and Humphreys, L. G., The effect of supporting and antagonistic voluntary instructions on conditioned discrimination. *J. Exper. Psychol.*, 1938, *22*:291-304.

Razran, G. H. S., Attitudinal control of human conditioning, *J. Psychol.*, 1936, *2*:327-337.

»»» 15 «««

EXPERIMENTAL STUDY OF HUMAN LEARNING

FOR over half a century, experimenters have pursued a patient search for the conditions of which the learner's performance is a function, and for the most efficient methods of measurement and analysis. They have found the number of significant variables large and the modes of interaction of these variables even more multiple and complex. The ever-growing body of experimental data has led only very slowly to the emergence of broad generalizations and unification in terms of systematic theory. There exists today a solid body of factual knowledge concerning the conditions which influence the speed and efficiency of verbal learning. A comprehensive theory of the processes mediating such learning is still in the making. In this chapter, our main emphasis will be on the experimental analysis of significant variables in verbal learning. Only occasionally shall we catch a glimpse of a unifying theoretical trend.

Methods of Practice

Serial Learning. Much verbal learning is *serial* in nature. The material which the learner must master consists of a series of items—nonsense syllables, nonsense consonants, digits, meaningful words. At the beginning of the experiment, the subject may or may not be acquainted with the individual items. Meaningful words and numbers are part of his daily repertory; most of the nonsense materials will probably be new to him. Whatever his acquaintance with the component items, his specific task is to connect them with each other, to form associations among them so that the disjointed parts come to form an organized pattern. Memorizing a list of nonsense syllables, a prose passage, or stanzas of poetry are all examples of serial learning. In the case of prose and poetry, the series is characterized by meaning, but in the case of nonsense syllables, the total

series has little more meaning than each component part. In both cases, however, learning depends on the formation of associations among the members of the series.[1] Much of the study of verbal learning, therefore, is concerned with the conditions leading to associations among the members of a series.

The members of a series are associated with each other as a result of practice. Several methods of stimulus presentation and practice have been developed in experiments on serial learning.

The Method of Complete Presentation. The total series is presented to the subject, and he is allowed to read it and explore it at his own speed. A time limit is usually set, but within the allotted time the learner's behavior is not controlled and he is free to linger over some parts, skip lightly over others. Nor can the experimenter be sure at what point the subject begins to rehearse the material and to recite to himself. When the experimenter finally tests the subject's retention, he may find it difficult to interpret the result for he has no information about the ways in which the subject attacked his task or about the temporal course of his learning. The main advantage of the method of complete presentation is its similarity to practical learning situations. When we study a lesson or prepare for an examination, the procedure which we use often closely resembles the method of complete presentation.

The Anticipation Method. The members of the series are presented one by one, at a regular rate, usually through the window of a memory drum. The full series is presented once, and thereafter the subject is instructed to *anticipate* each item before it appears. The correct item is presented whether or not the subject has attempted to anticipate it. If he does attempt an anticipation, the appearance of the correct item serves to verify or correct his response. If he fails to anticipate, he is prompted by the presentation of the correct item. Thus, each presentation of the series is both a learning trial and a test of retention. This procedure is usually continued until the subject has reached the criterion set by the experimenter. More often than not, this criterion calls for one errorless series of anticipations, but more exacting criteria, such as three successive

[1] We remind the reader of the purely descriptive sense in which the term association is used. The term does not connote the formation of neural connections or similar explanatory hypotheses.

errorless series of anticipations, are sometimes used. The greatest advantage of the method of anticipation is that it allows the experimenter to obtain a full picture of the temporal course of learning, both for the series as a whole and for each individual item. We can see at what point any given item is correctly anticipated for the first time and how stable the retention for the item is, i.e., how often it is learned, forgotten, and relearned. We can gauge the amount of improvement which results from a test trial at various stages of learning. The method is ideally suited for the study of serial learning, for it is in terms of their serial position that the subject must connect the items with each other.

The Method of Paired Associates. Anyone who has tried to memorize a list of foreign words and their English equivalents is familiar with this method. On the first trial, paired items (e.g., two nonsense syllables, a syllable and a number, two meaningful words, or a meaningful word and a number) are presented to the subject. On all subsequent trials, the first member of the pair is presented to the subject, and he attempts to recall the second. The subject is given a limited amount of time in which to supply the missing member of the pair, after which the correct item is presented to him. The method thus is a special case of the method of anticipation, and again each trial is both a learning trial and a test of retention. Again, the temporal course of learning can be followed in detail. The emphasis here is on connection between the members of a pair (each of which may be considered a series of two items) rather than on the connection of a long series of individual items. It is customary, therefore, to present the pairs in a different random order on each trial in order to prevent the subject from learning the second members of the pairs as a series of responses independent of the first members.

The three methods discussed here represent the basic procedures used in the practice of verbal materials. They are used for serial learning as well as for learning that is not primarily serial in nature. In experimental practice, departures from these standard models are frequently introduced, and features of several methods may be combined in order to find the answer to a particular problem. As in the case of psychophysical methods (see pp. 14-15), these procedures must not be considered as cut and dried, inflexible prescriptions but rather as models which the experimenter is free to

vary and adapt according to his needs. Two aspects of the procedure which are often varied are the order of presentation of the stimulus items and the criterion to which learning is carried.

Uniform vs. Random Order of Presentation. On the practice trials, the items of the series may be presented either in uniform or in random order. If the order is uniform, the sequence of items remains unchanged from trial to trial. If the order is random, an entirely different sequence is used on every trial. The subject's task and the learning process are radically affected by the order of presentation. If the order is uniform, the subject's effort is directed toward learning the items in their proper serial order, and the strength of association among the various items will depend largely on their distance from each other in the series. If the order is random, the learner must put his main emphasis on the items themselves, often disregarding serial position in any particular trial. As we have seen, the method of anticipation necessarily implies a uniform order of presentation; the method of paired associates favors a random order of pairs. Either order may be used with the method of complete presentation. In designing an experiment, we must always ask whether we are interested in studying the effects of serial position on the learning of the individual items or whether we wish to eliminate the effects of position in order to study another variable independently of such effects. For example, if we were concerned with the question of whether "emotional" words are learned more or less rapidly than "neutral" ones, we would present the items in random order and in this way equalize the effects of serial position for both types of items. On the other hand, if we wish to compare the ease of learning items located at the beginning, center, and end of a list, we would keep the order of presentation constant from trial to trial.

Criterion to Which Learning Is Carried. The period of practice in a learning experiment may be extended indefinitely. There is no one point at which it would be possible to state with any assurance that learning has ceased. Even after a learner has mastered a list of items—i.e., when he is able to recite it without error—he may benefit from further practice. His performance may become smoother, more automatic. The more he practices the list, the less quickly will he forget it. Since the period of learning and improvement is indefinite, it is necessary to decide, for a given experi-

ment, on a *criterion of performance* to which learning is to be carried. One frequently used criterion, which we have already mentioned, is *complete mastery*, i.e., one errorless recall of the material. If practice is continued beyond that point to, say, three or five errorless repetitions, overlearning takes place. It would be difficult to state a general rule concerning the best criterion in an experiment on verbal learning. The optimal criterion will vary from experiment to experiment. In many experiments, the criterion of complete mastery has been found useful, but less exacting criteria, such as correct repetitions of 75 percent of the items, have also been used successfully. If degree of learning is one of the independent variables, several criteria may be used within one experimental design.

Instead of carrying learning to a criterion, it is possible to practice the material for a fixed number of trials and then test for retention. Such a procedure may often save time, especially in the performance of group experiments. It has, however, one serious disadvantage. At the end of a fixed number of trials, there will be considerable variability in the degree of mastery achieved by individual learners. Differences in learning ability are appreciable, even among subjects with fairly homogeneous educational and social backgrounds. Thus, with a constant amount of practice, the amount of learning is far from constant from subject to subject. Generalization about the effects of experimental variables becomes difficult when individual differences in strength of learning are uncontrolled. If all subjects are carried to the same criterion, their results become more comparable.

Mastery, as revealed by the ability to reproduce the material, provides, of course, only one of several possible criteria of learning, although it is the one most frequently used. Other criteria may be set, e.g., the ability to recognize a certain percentage of items correctly, or a minimum speed of performance. The choice of criterion depends on the particular method used to measure the strength of learning.

The Basic Variables in Learning Experiments

In any experimental situation, the learner's performance (however measured) is a joint function of many conditions. As we have emphasized, one of the most striking facts revealed by investigations

in this field is the multiplicity and complexity of the variables which significantly influence learning. To evaluate the effects of these variables, it is desirable to vary one or a few at a time, holding all other relevant factors as constant as possible. Three important sets of variables have been investigated under conditions of controlled experimentation: (1) the nature of the materials or activities learned; (2) the conditions of practice under which learning takes place; and (3) the personal characteristics of the subjects, such as their age, sex, and intelligence. Measurable performance depends on *what* is learned, *how* it is learned, and *who* learns it. We shall briefly summarize the most important methodological and experimental findings under each of these three headings.

Performance as a Function of What Is Learned

There are few human activities which do not benefit from learning, but the effects are by no means uniform for all activities. Different activities are learned at different rates. Optimal conditions of practice also vary from activity to activity. From the large amount of information bearing on the problem of learning as a function of what is learned, we shall select a few generalizations concerning verbal materials.

Dimensional Analysis. Verbal materials may differ from each other in many ways. A systematic comparison of such materials can best be carried out through *dimensional analysis*.[2] We must begin by stating the characteristics or dimensions, with respect to which verbal materials vary from each other. Two sets of verbal materials may differ in one characteristic or dimension and be alike in another or, as frequently as not, they may differ from each other in several dimensions. The effect of variation in each of the dimensions can then be studied, holding constant (experimentally and/or statistically) variations in other dimensions. We shall consider three dimensions of variation in verbal material which have been isolated experimentally: (1) meaningfulness, (2) affectivity, and (3) amount.

[2] The need for dimensional analysis was stressed by J. A. McGeoch, *The psychology of human learning*, New York: Longmans, Green and Co., 1942. This treatise contains a most thorough and comprehensive analysis of the determinants of human learning.

The Dimension of Meaning. Verbal materials may or may not have standard dictionary meaning. As we have seen, the nonsense syllable was invented with the express purpose of eliminating meaning in order to study "pure" learning and retention, unencumbered by uncontrolled associations introduced by meaning. It is now clear that meaning is a question of degree. Few nonsense items are completely meaningless, and even if they lack meaning at the beginning of an experiment, the subject soon succeeds in making them more or less meaningful by invoking associations and similarities, however far-fetched and remote. The advantage of nonsense materials lies primarily in the fact that they allow the use of materials of fairly *uniform* difficulty. A list of nonsense syllables can be constructed so that its members are reasonably equal in difficulty and susceptibility to meaningful associations. Individual subjects, moreover, may be assumed to have fairly equal degrees of acquaintance with nonsense material (provided previous service in learning experiments is held constant), whereas experience with meaningful words varies much more widely from individual to individual.

The difference in degree of meaning between nonsense materials and standard dictionary words is reflected in the speed with which words are learned. There is general agreement that meaningful items are learned significantly faster than nonsense materials. This generalization holds whether the nonsense materials be syllables, consonants, or groups of digits, and the meaningful materials be discrete words, prose passages, or stanzas of poetry.

Within the two broad classes of nonsense and meaningful materials, individual items vary continuously in degree of meaning. Some nonsense words suggest meanings and associations more readily than others (cf. p. 279). Other things being equal, the greater the "association value" of a nonsense item, i.e., the more frequently it reminds subjects of meaningful words, the more quickly it is learned. Meaningful materials themselves vary in degree of meaning and consequently in ease of learning. A series of discrete words is learned more slowly than a connected passage. A prose passage, in turn, is likely to be learned less quickly than a comparable amount of poetry which has the additional advantage of rhyme and rhythm.

Meaning facilitates learning because it makes it easier to connect individual items with each other, to bring to bear on the new learn-

ing task previous learning and previous experience. Especially with serial materials, rate of learning depends largely on the ease with which the subject can organize and group the items. Meaning is an important aid in achieving such organization.

The Dimension of Affectivity. Meaning facilitates learning by allowing the subject to utilize past learning in the formation of new associations. The kinds of past experiences which may function in the acquisition of new materials are, of course, many and varied. Past experience with the learning material may, for example, be characterized by differences in affectivity or "emotional tone." Meaningful words may be pleasant, unpleasant, or neutral, being associated to varying degrees with emotional experiences in the past. There have been many studies of rote learning as a function of the affective characteristics of the material. The results have not always been univocal, but, in general, pleasant stimulus items have shown quickest learning, with unpleasant and neutral ones following in that order. A serious methodological problem is the difficulty of arriving at valid independent classifications of learning materials as pleasant, unpleasant, and neutral. *A priori* classifications merely reflect the experimenter's judgment and may have little or no validity for some of his subjects. Frequently, experimenters have asked the subjects themselves to judge the materials some time before or after learning. As far as the subjects' judgments are reliable and valid, they may legitimately be used in studying the correlation between affectivity and rate of learning. Affective judgments, however, are difficult to validate (cf. p. 231). Because of the very fact that emotional responses are involved, the subject himself may be a poor judge of the affective nature of the material. Affectivity is not a variable which can be easily measured or varied experimentally.

The dimension of affectivity is not limited to pleasantness-unpleasantness. Verbal material may evoke a variety of attitudes in the learner, and such attitudes may be significant determinants of the rate of learning. It has been shown, for example, that materials congruent with the learner's political preferences may be learned more quickly and forgotten more slowly than materials to which he is opposed. In other experiments, sex differences in learning performance have been traced to differences in attitude or interest in the material learned. Findings such as these do not as yet have the

status of well-established generalizations, for the study of learning in relation to attitudes and interests is still very much in its beginnings This area of investigation is, however, of great potential significance because it helps to extend the horizon of learning experiments to problems of social psychology and personality.

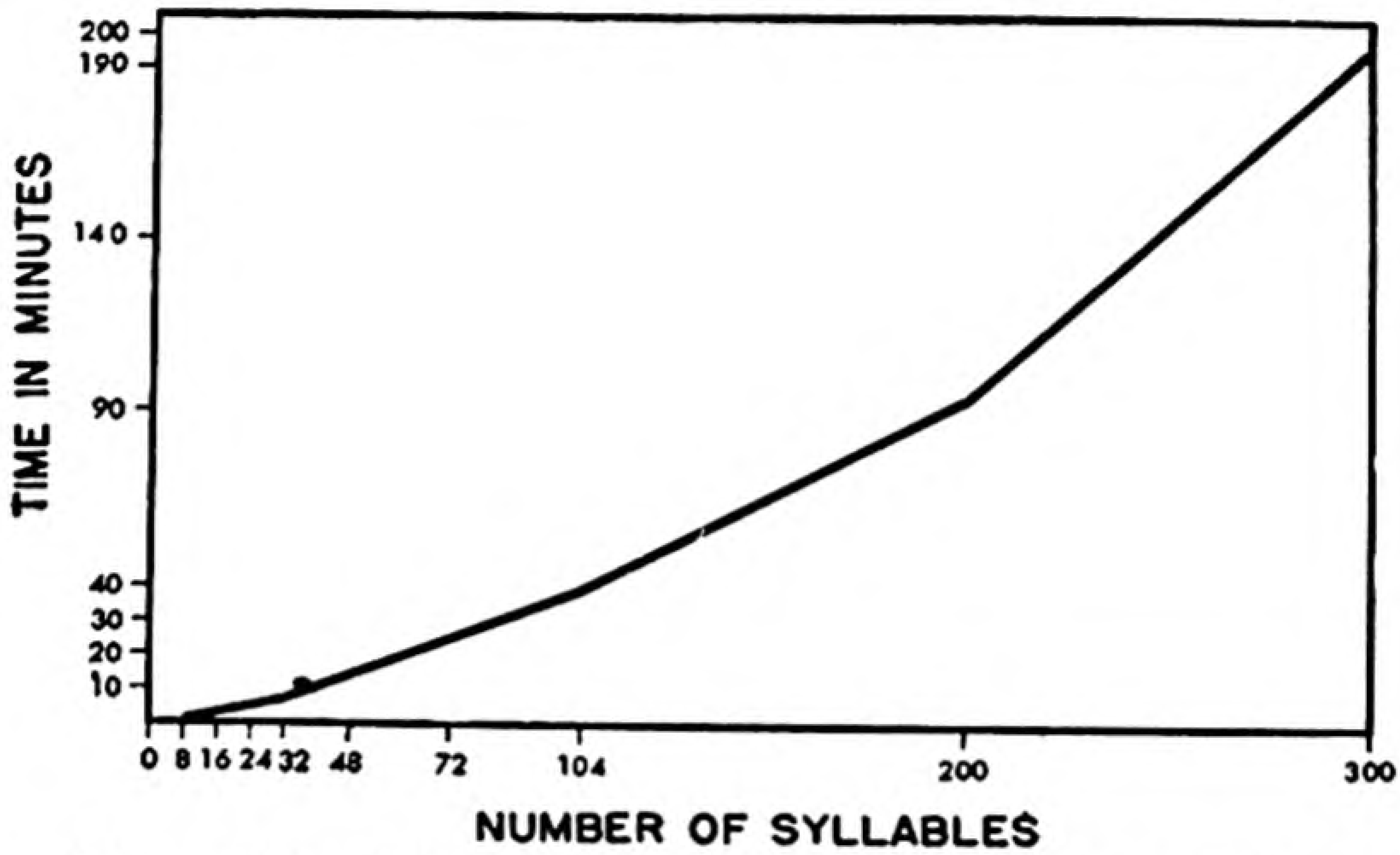


FIG. 83. The Effect of Length of List on Difficulty of Learning. In this figure, the amount of time required to learn a list of nonsense syllables is plotted against the number of syllables in the list. Clearly, the longer the list, the longer it takes to learn it. In addition, the relationship is not linear, i.e., the longer the list, the more time is required per item learned. (From J. A. McGeoch, *The psychology of human learning*, 1942, p. 178, by permission of Longmans, Green & Co., Inc. After D. O. Lyon, *Memory and the learning process*, 1917, by permission of Warwick & York, Inc.)

The Dimension of Amount Learned. Whatever the nature of the individual items in a learning task, the learner must establish connections among them, group them, and organize them so that he may finally perform the whole task (or meet some other criterion set by the experimenter). *Amount* to be learned is one of the dimensions along which verbal materials vary, quite apart from their specific nature. How is rate of learning affected by the magnitude of the subject's total task? As far as sheer amount of time to reach criterion is concerned, a long task must necessarily take more time

than a short one. If the items are presented or read at a constant speed, each presentation of a long list will occupy more time than that of a short list. We may further ask, however, whether the *amount of time per item learned* remains constant or changes as the length of a task is varied. The experimental evidence clearly shows that the amount of time per item does become greater as the length of the learning task is increased. Fig. 83 illustrates this point. As the number of syllables in a list increases, the total time needed to learn them increases disproportionately. The curve is not a straight line but shows positive acceleration. The longer the list, the longer we have to spend on each item in order to reach a fixed criterion such as complete mastery. Although the relationship depicted in Fig. 83 is clearly not linear, the amount of positive acceleration is not very great. Over a wide range, the curve closely approximates a straight line. The disproportionate increase in time per list becomes really serious only when the list exceeds 200 items.

Disproportionate increases in learning time with increased amounts of material have been found with a variety of activities and materials, and the relationship may be considered a general one. The degree of disproportionality may, however, vary from one learning situation to another. Sheer rote learning shows the effect to a greater extent than does "learning for understanding." For practiced subjects, a longer task means less of an increase in difficulty than for untrained ones. Variations such as these suggest that the increase in time per item is due to the fact that the learner finds it more and more difficult to group or organize his responses as the amount of material becomes larger and larger. In a long list, the individual items are apt to interfere with each other, increasing the number of incorrect associations. We should expect this effect to be most serious in rote learning where the exact sequence of the individual items is important. An experienced learner will be more skillful than an untrained subject in overcoming such interference effects and in using various aids in organizing and grouping the material.

The Effects of Serial Position

The Serial-Position Curve. When a list of items is learned in a definite order (e.g., by the anticipation method), the speed with

which any given member of the series is learned varies with its position in the series. In general, the first item in the series is learned fastest; thereafter, speed of learning decreases as a function of serial position, reaches a minimum near the middle of the series,

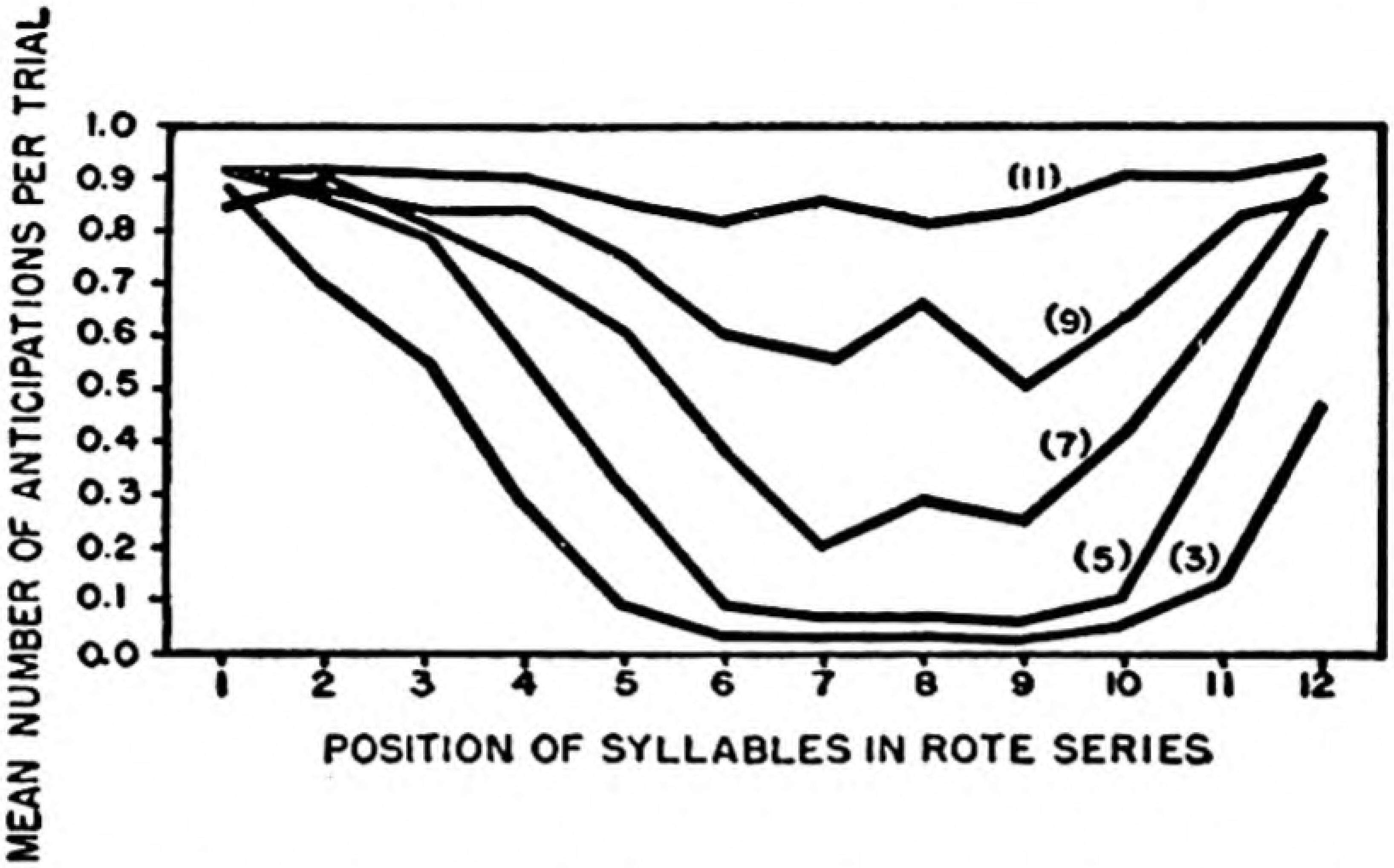


FIG. 84. A Set of Serial Position Curves. The number of correct anticipations per trial is plotted against the position of the syllable in the list. The numbers on the curves refer to the criterion met by the subject on the trial for which the curve is plotted. In the early stages of learning there is a striking sag toward the center of the list. As learning approaches completion, the curve necessarily levels out more and more. (From J. A. McGeoch, *The psychology of human learning,* 1942, p. 98, by permission of Longmans, Green & Co., Inc. After L. B. Ward, Reminiscence and rote learning, *Psychol. Monogr.,* 1937, *49,* No. 220, p. 35, by permission of the journal and the American Psychological Association.)

and then rises again. The last item is learned considerably faster than the middle ones though usually not as well as the first one. Fig. 84 shows a set of typical, "bow-shaped," serial-position curves. In Fig. 84, the mean number of correct anticipations per trial is plotted against the serial position of the item. As these curves show, the effects of serial position are greatest in the early stages of practice. When only three or five successful anticipations are made on

a given trial, only a very small proportion of the correct responses occurs in the middle part of the list. As learning progresses, more and more of the items in the middle of the list are learned, and, as a result, the differences among the serial positions diminish and the curve is flattened out. If learning is carried to a criterion of full mastery, the curve for the final trial is, by definition, a straight line, since all the items are recalled correctly on the criterion trial. There are, however, always more failures in the middle positions than at the ends of the series before mastery is reached.

Intraserial Associations. The effects of serial position are frequently ascribed to the action of remote associations. When a series

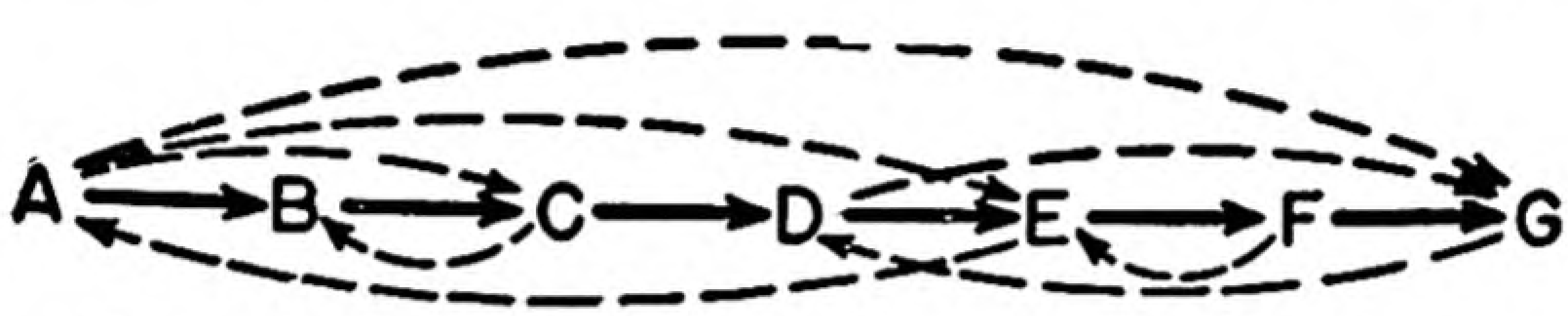


FIG. 85. Immediate and Remote Associations Formed Between the Items in a Learning Series. The letters *A* through *G* represent items in the list, e.g., nonsense syllables. The heavy arrows denote immediate associations. The dotted arrows represent remote associations. (From J. A. McGeoch, *The psychology of human learning,* 1942, p. 70, by permission of Longmans, Green & Co., Inc.)

of items is learned, associations are formed not only between adjacent members of the list but also among items which are more than one step removed from each other. Such associations are called *remote.* Remote associations may be in the forward or backward direction. Before attempting an explanation of serial-position effects, let us consider the nature of such remote associations.

Consider Fig. 85. The letters *A* through *G* represent items in a list. The heavy arrows represent associations formed between adjacent items, the dotted arrows represent remote associations. As the subject practices the list, he not only connects *A* with *B, B* with *C,* and so on, but also *A* with *C, A* with *D, E* with *A, G* with *D,* and so on. Backward remote associations are much less frequent than forward ones. The existence of remote associations has been inferred from a variety of experimental data. The *method of derived lists* has provided one of the main lines of evidence. After the subject has

learned the list, we can reshuffle the items (for example, in the order *A C E G B D F*) and have the subject learn the new sequence. He will learn the new or "derived" list faster than the original list. The fact that he learns the derived list faster suggests that remote associations were formed during the original learning which facilitated mastery of the new task.

Evidence for the formation of remote associations also comes from an analysis of the errors which the subject makes while practicing a list of items. Not infrequently we note "anticipatory" errors. In recalling the items, the subject will sometimes skip one or more intermediate steps and connect remote items with each other. Thus, after recalling *A*, he may reproduce *C* or *D*, or even *E*, *F*, or *G*. Such errors are called anticipatory because the items reproduced, though correct, are given too early in the sequence. Without postulating remote associations, such errors would be difficult to explain. While anticipatory errors may be ascribed to remote associations in the forward direction, "perseverative" errors may be due to remote associations in the backward direction. A perseverative error occurs when a previous correct response is repeated later on in the series. For example, *C* may be correctly recalled following *B*, and then be repeated following *E*. In general, the further two items are separated from each other, the less likely are they to be connected by anticipatory or perseverative error. Both anticipatory and perseverative errors decrease, of course, as learning progresses toward the criterion. Progress of learning is thus characterized by the decrease in inhibition due to remote associations, and the strengthening of immediate associations.

The effects of serial position can be ascribed to the action of remote associations. Let us consider only remote forward associations which independent evidence shows to be much more frequent and much stronger than backward ones. As the series is practiced, a large number of such remote associations is formed. As Fig. 86 shows, *the number of remote associations spanning a given item depends on its serial position.* More remote forward associations span items in the middle of the series than items toward the ends of the series. We have seen that reaching a criterion of complete mastery (say, by the method of anticipation) depends on strengthening of the immediate associations relative to the remote associations, for only on this condition can the serial order be re-

produced correctly. Since the items in the middle of the series are spanned by the largest number of remote associations, such items will be subject to the greatest amount of inhibition in the course of

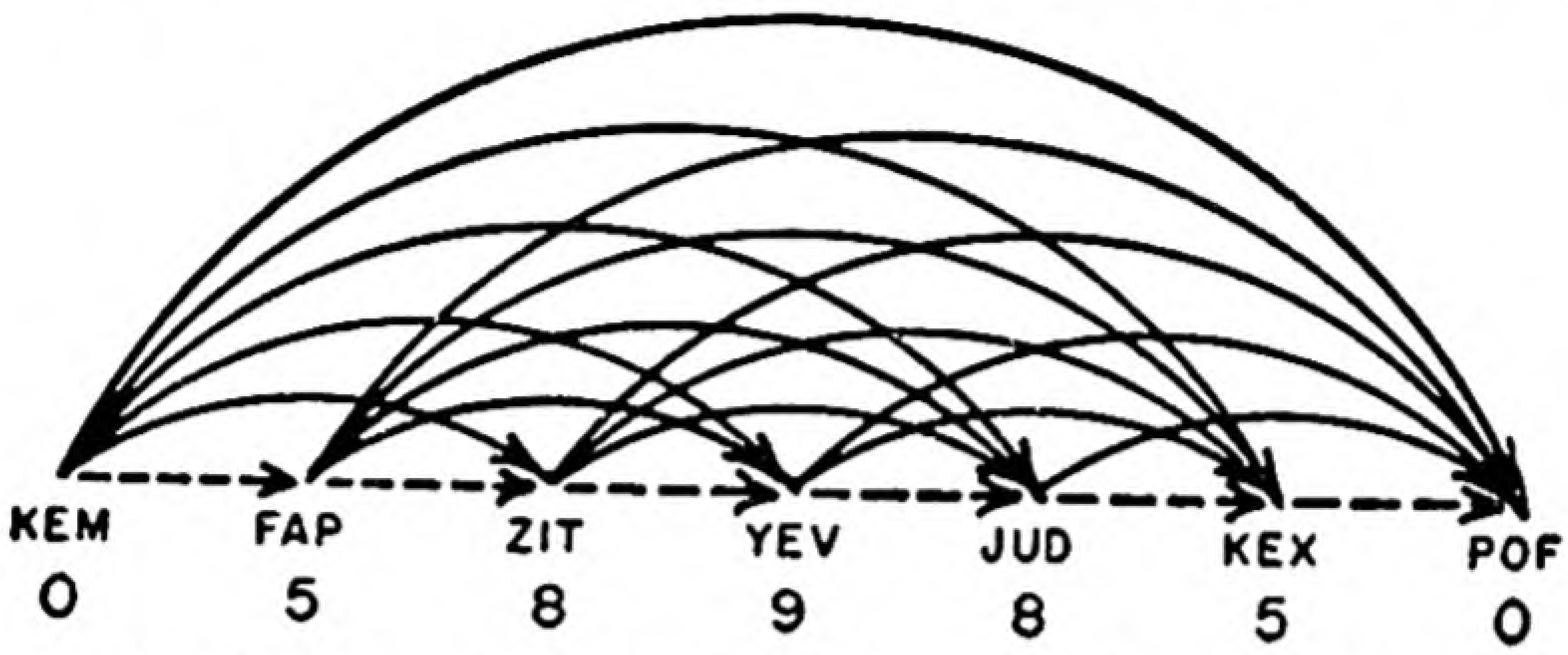


FIG. 86. A Schematic Diagram Showing the Immediate and Remote Associations Which Are Believed to Span the Items in a Learning Series. The broken arrows represent associations between adjacent items. The curved arrows represent remote forward associations. The numbers under the nonsense syllables indicate the number of forward remote associations presumably spanning that item. The items in the middle of the list are spanned by the largest number of associations. (From J. A. McGeoch, *The psychology of human learning,* 1942, p. 109, by permission of Longmans, Green & Co., Inc. After C. L. Hull, The conflicting psychologies of learning—a way out, *Psychol. Rev.,* 1935, *42*:502, by permission of the journal and the American Psychological Association.)

practice, and it will take a relatively long time to overcome such inhibition and to strengthen immediate associations sufficiently. Thus, serial-position effects may be ascribed to the inhibitory action of remote associations which must be overcome before complete mastery can be obtained.

In our discussion, we have been speaking of "remote" associations. Remoteness should be understood to refer only to position in the series. In the last analysis, remote associations may turn out to be no less immediate than associations between adjacent items in the series. Whatever neural counterparts the perception of a syllable has, they probably persist for a small finite period of time after the syllable has been presented, and it is this persisting "trace" which may immediately be associated with a remote item. Thus,

turning to Fig. 86, the aftereffect or trace of *KEM* may still be active when *ZIT* is presented. Thus, the trace of *KEM* is associated with *ZIT*, and a remote association is formed. Such speculations serve to emphasize that immediate and remote associations are probably not qualitatively different.

Performance as a Function of How Learning Proceeds

No matter what the material or activity learned, the speed and efficiency of the subject's performance are greatly influenced by the conditions under which he practices, his motivation, and his mode of attack. In this section we shall consider a selected group of conditions of practice which significantly influence the learner's performance.

Set and Motivation

In the experimental analysis of verbal learning, no less than in the analysis of all other behavior, the problem of motivation must be considered. There is a twofold problem here: (1) with what motivation does the learner approach his task; and (2) how do changes in his motivation, his successes and failures, affect the progress of his learning? Turning to the first of these problems, we find that the primary motivational condition is the subject's intent to learn.

Intent to Learn. When a subject comes to the laboratory to participate in a learning experiment, he comes with the intent to learn. The motivation for this behavior is complex and its nature usually has not been investigated. Curiosity, interest in the subject matter, and the desire to please the experimenter (or the need to placate him if he happens to be an instructor) all play a part. It is the intent to learn which constitutes the crucial difference between passive exposure to stimulus objects and the active effort of a learner. Without intent to learn, little learning takes place. We may be exposed to stimuli literally hundreds of times and fail to learn them or learn them very poorly if we never had the intention to remember them.[3] In most experiments on verbal learning, it is

[3] We shall return to this problem later, under the heading of incidental learning.

sufficient to instruct the subject in order to create in him adequate motivation or intention to learn. The experimenter working with human learners is, in this regard, more fortunate than the animal psychologist who has to use hunger, thirst, and shock to create the proper motivation in his subjects.

The Role of Set. Intent to learn is often described as a *set* to learn. Though widely used in experimental psychology, the concept of set has sometimes lacked clear definition. In its broadest sense, *set* denotes a readiness to respond to stimulus objects in a selective way, to respond to some objects and specific characteristics rather than others. By its set, the organism is selectively tuned to events in the environment. Thus, in reading a prose passage, our set may lead us to read it for understanding, for aesthetic enjoyment, or for the purpose of reproducing it verbatim. Frequently two or more such sets may be operative at the same time.

In addition to the general set to learn, the subject usually has a more specific set toward the material or activity which he practices. He may be set to learn a series of items in their exact sequence, to learn them without regard to sequence so as to be able to recognize them later, or to learn simply with a view to answering questions about the content of the material. Usually this specific set is established by instructions and the nature of the experimental procedure. In the method of anticipation, for example, instructions focus the subject's attention on the serial order of the items, and the procedure of exposing the items on a memory drum and requiring the subject to name them before they appear strengthens the set for serial association.

Learning proceeds with greatest efficiency if the subject's set is sharply focused on that aspect of the material in terms of which learning is defined and measured. In serial learning, a sustained set for serial position is important; in learning for recognition, an equally sustained set for the "content" of the material, quite apart from the sequence in which it is presented, is desirable. Sets are selective in what they emphasize and also in what they exclude.

In experiments on verbal learning, sets are usually established by the experimenter's instructions to the subject. It would be dangerous to assume, however, that in the absence of specific instructions the subject operates without any set. Sets are not merely the

result of instructions; they often represent well-established habits of perceiving and responding. We rarely read, for example, with a completely passive attitude. Whether or not we have been so instructed, we are set to understand and perhaps to reproduce what we have read. Instructions do not, then, impose a set on an otherwise "setless" organism; rather, they control and specify the set.

Incidental Learning. It is precisely because sets to learn seem to be ubiquitous, even in the absence of specific instructions, that "incidental learning" is so difficult to interpret. Learning which occurs without specific instructions to learn has often been designated as *incidental*. In a typical experiment on incidental learning, the subject may be instructed to learn a list of words printed on a sheet of paper and may then be tested for his memory of such physical details as the color of the paper, designs on the margin of the sheet, etc. Since he was not specifically instructed to learn these physical details, whatever memory he shows for them is considered incidental. Various other procedures have been used to test incidental learning, but they all have two conditions in common: (1) the materials must be presented in such a way that the sense organs are stimulated by the incidental items; and (2) the subject is *not* instructed to learn these items and his attention is directed to other parts or characteristics of the material.

The experimental evidence is clear: incidental learning does take place although performance is never so good as under instructions to learn. The occurrence of so-called incidental learning, however, does not prove that learning can ever take place in the absence of a set to learn. Subjects may instruct themselves to learn even when the experimenter fails to do so, and there is good evidence that they, in fact, do so. A circumscribed set suggested by the experimenter's instruction may be generalized to other incidental items not covered by the instructions. Especially in an experimental situation, in which a subject is alert and motivated to respond actively to the task before him, generalization of a learning set to extraneous features of the situation may easily take place. The weight of the evidence points to the conclusion that learning is never truly incidental in the sense of being undetermined or haphazard.

Let us now consider motivational changes which take place in

the course of the learning process, notably the effects of successes (rewards) and failures (punishments).

The Effects of Success and Failure. Consider a subject in an anticipation experiment. He tries to pronounce a series of nonsense syllables before they appear in the window of the memory drum. As soon as he has made his anticipation, the correct syllable appears. Sometimes his anticipation proves right, but frequently it is wrong, especially in the early stages of learning. If his anticipation was correct, he is more likely than not to repeat it on the next trial. If his anticipation was wrong, he will probably not repeat it on the next trial and eventually make the correct response. Success and failure are extremely important determinants of progressive changes in behavior. This is true for verbal learning as well as for much other learning, both of animals and of human subjects.

Successful responses are strengthened; unsuccessful ones are not and eventually drop out. Much of modern learning theory is based on this general principle—the principle of reinforcement[4] or "law of effect." The principle has found its most striking demonstration in animal experiments. If turning right in a maze leads to food while turning left does not, a hungry animal will learn to choose the right path whenever he is put in the choice situation. The successful response, i.e., the response leading to reinforcement is strengthened at the expense of the response which fails to lead to reinforcement.

Reinforcement by reward probably plays an important part in verbal learning but its demonstration has not always been easy. One particular experimental situation has probably been used more than any other to show the efficacy of rewards. The procedure is essentially an adaptation of the method of paired associates. The experimenter presents a long list of words to the subject who is required to guess a number from one to ten in response to each of the words. The experimenter arbitrarily calls a few of these numbers "Right" and fails to reward the others or calls them "Wrong." It has been shown again and again that the numbers called "Right" are learned faster (remembered better) than those which are not so reinforced. At the same time, it was found that punishment by "Wrong" is not an efficient means of eliminating false responses.

[4] For further discussion of the principle of reinforcement, see Chapter 14.

Certainly, punishment by "Wrong" weakens a response much less than reward by "Right" strengthens it. An announcement of "Wrong" usually serves to make responses more variable but does not necessarily reduce the probability of repetition of the false response. Whatever influence punishment exerts is indirect: by making behavior more variable, it increases the opportunity for the correct response to be given and strengthened by reward.

The question has often been raised whether announcements of "Right" and "Wrong" can legitimately be considered rewards and punishments. Are they not simply ways of conveying information to the subject? The question is difficult to answer and largely a matter of definition. The important thing is that the consequences of a response (announcement of "Right" or "Wrong") significantly influence further responses.

In verbal learning experiments, reinforcement (reward) need not necessarily originate with the experimenter. The learning situation itself may be a source of reinforcements. Anticipating a syllable correctly and in time and reciting a list without error are reinforcing states of affairs even if no announcement of "Right" or other approval comes from the experimenter. The quick modification of responses and the selective elimination of errors which are so characteristic of verbal learning greatly favor an explanation of the learning in terms of reinforcement.

Knowledge of Results and Active Recitation. Continuous with the effects of success and failure is the influence which knowledge of results exercises on learning performance. If the subject is kept informed of his progress—e.g., if he is given his score on successive trials—his learning is faster and better than if he is kept in ignorance of the results throughout the practice period. Knowledge of results probably serves a double purpose: (1) it serves as "reward" and "punishment," especially in relation to the goals which the subject sets for himself; and (2) it serves to guide the learner's efforts, enabling him to evaluate the efficiency of his attack on the problem by the results which he achieves.

Even if no information is supplied by the experimenter, the subject can achieve at least some knowledge of results by active recitation of the material. He can readily discover how far he is still from

his goal, where his strength and his weakness are. Active recitation during practice, supplemented by prompting, has been shown experimentally to be a most effective method of practice. The earlier active recitation is introduced, the faster and the more efficient is the course of learning. It is efficient to allot a considerable proportion of the total learning time to active recitation, even as much as 4/5 of the total practice period. As we have seen, recitation provides knowledge of results and helps to direct the learner's efforts. In addition, active recitation is a most suitable method of practice in relation to the learner's goal. The learner is striving to acquire new responses and response patterns. The earlier he actually attempts to perform these responses by active recitation, the more directly does he drive at his eventual goal.

Distribution of Work

Speed and efficiency of learning depend not only on the nature of the activity and the subject's motivation but also on his work methods or modes of attack on the problem. We are dealing here with the same type of problem that you know under the heading of "study habits." What is the best way to go about learning material for a test? It is necessary to state at the outset that there is no simple, all-purpose answer to this question. The psychologist must sound the familiar warning that optimal methods of practice vary from situation to situation and, indeed, from individual to individual. On the basis of available experimental evidence, certain generalizations can nevertheless be made.

Whole vs. Part Methods of Study. Whenever the material to be learned consists of a series of items, it may be practiced either as a whole or in separate parts. In memorizing a poem, for example, one may try to practice the whole poem or break it up into stanzas. Practice may be by the *whole method* or the *part method.*

A great deal of experimental work has been devoted to an evaluation of whole vs. part methods of practice, and the results have not always been free of contradiction. Superior efficiency has been claimed for each of the various methods. These contradictions are probably due to the fact that neither length nor nature of the materials has been standardized, which makes comparison of different

experiments difficult, if not impossible. The balance of the evidence, however, favors the whole method, especially for practiced learners and meaningful materials.

A consideration of the time relationships involved may help to explain the frequent superiority of the whole method and also some of its failures to surpass the part method. We already know that time-per-item increases with increases in amount of material. Time-per-item will, therefore, generally be less when the part method is used. However, once the separate parts are learned, a considerable amount of time must be spent combining the parts into a whole, and the advantage of shorter time-per-unit, inherent in the part method, may be lost. Such is indeed frequently, but not always, the case.

Comparison of the whole and part methods involves more than the algebraic sums of the times spent in learning by the two procedures. They represent different ways of organizing the material. If learning is by the whole method, the total pattern of the task is before the learner at all times, and individual parts are gradually fitted in. If learning is by the part method, the organization of the total pattern of responses is delayed.

The more closely organized and meaningful the material is, the more important it becomes to apprehend the total pattern at an early stage of the learning. It has, indeed, been frequently found that the whole method is especially suitable for complex and meaningful materials. Conversely, the part method, which allows the learner to concentrate on a few specific responses at a time, has sometimes proved superior in the rote learning of disparate items.

It is not surprising that the whole method yields its best results with experienced learners. Practiced subjects are more skillful at handling large amounts of material and organizing them. They also save less time-per-item using the part method than do inexperienced learners. The ability to achieve superior results by the whole method is itself a result of learning.

Distribution of Practice. Just as the learning *material* may be attacked as a whole or in parts, so learning *time* may either be spent continuously or spaced out over periods of varying lengths. Practice may be *massed* or *distributed* in time. Suppose we have *n* minutes at our disposal to devote to learning a task. Under conditions

of massed practice, we would use the entire *n* minutes in one continuous learning session. Under conditions of distributed practice, we should practice for a fraction of *n* minutes, take a rest period (where rest is defined as an activity unrelated to the learning task),

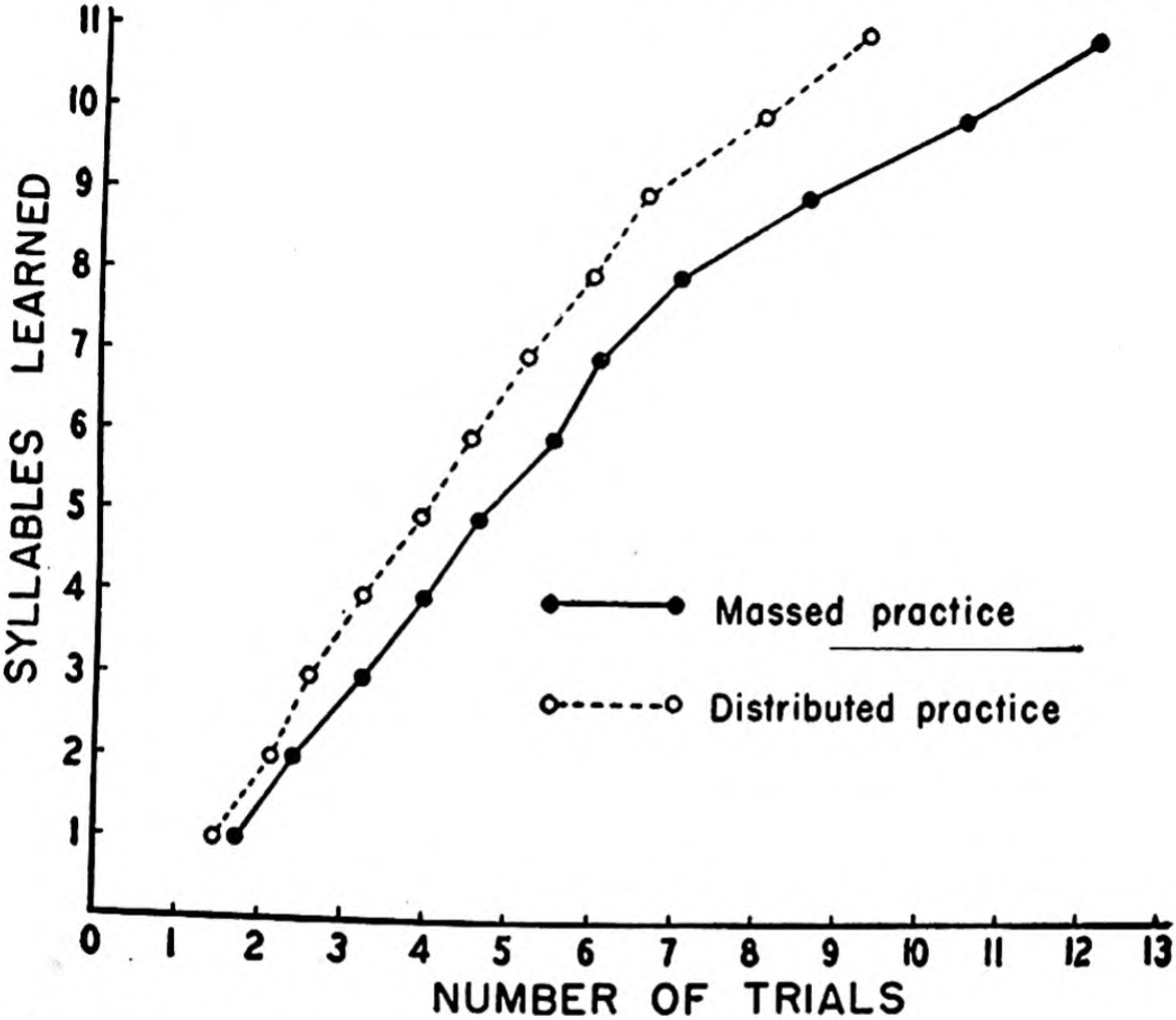


FIG. 87. Comparison of Massed and Distributed Practice. A series of syllables is learned faster under distributed than massed practice. (From C. I. Hovland, Experimental studies in rote learning theory: V. Comparison of distribution of practice in serial and paired associate learning, *J. Exper. Psychol.*, 1939, 25:625, by permission of the journal and the American Psychological Association.)

practice for another fraction of *n* minutes, interpolate another rest, and so on, until the allotted time has been used up.

The experimental evidence shows unequivocally that distribution of practice leads to faster learning and better retention. This finding has been confirmed again and again with a large variety of learning materials and under many different experimental conditions. Fig. 87

presents typical results, illustrating the faster rate of learning under distributed as compared with massed practice.

The extent to which practice is distributed may, of course, vary. Rest periods of different lengths may be interpolated between successive practice periods. The practice periods may themselves vary in length: they may be as short as one trial or comprise a long block of trials. The advantage gained by distribution of practice depends jointly on these two variables: duration of individual practice periods and duration of interpolated rest intervals. For a given activity, there is usually an optimal combination of practice periods and rest intervals which results in maximum learning efficiency. Other things being equal, rather short practice periods, followed by short rest intervals, are frequently found to be most effective and most practical. The longer the individual practice periods, the longer the rest intervals must be for optimal speed of learning and retention. Increasing the length of the rest interval beyond a certain optimal amount has, however, no further facilitating effect nor is such a procedure usually practical. Distribution should never be over periods so long as to lead to serious amounts of forgetting between successive practice periods.

Not all activities benefit equally from distribution of practice. Materials which are very easy or very short may be profitably learned at one sitting, i.e., by massed practice. The learning and retention of long and difficult materials, on the other hand, are greatly facilitated by distribution. The task of learning a list of nonsense syllables, for example, is shortened considerably by distribution. The beneficial effects of distribution have been demonstrated with a wide range of activities including perceptual-motor tasks as well as verbal material, both nonsense and meaningful.

We may learn an activity under conditions of distribution from beginning to end, or we may work with distribution part of the time and with massed practice part of the time. We may also vary the length of the rest interval at different stages of learning. No one clear-cut generalization about the effects of such variations is possible, and the optimal sequence of intervals apparently varies considerably with the nature of the activity learned. The following general considerations should be borne in mind: when stereotypy of behavior is undesirable, as for example, in the discovery of new

relations in problem solving, distribution is beneficial. When it is important to fixate a newly learned response before it is lost, massed practice may be advantageous. In the learning of a complex activity, both the discovery of new relations and the "stamping-in" of responses are necessary. A judicious sequence of distributed and

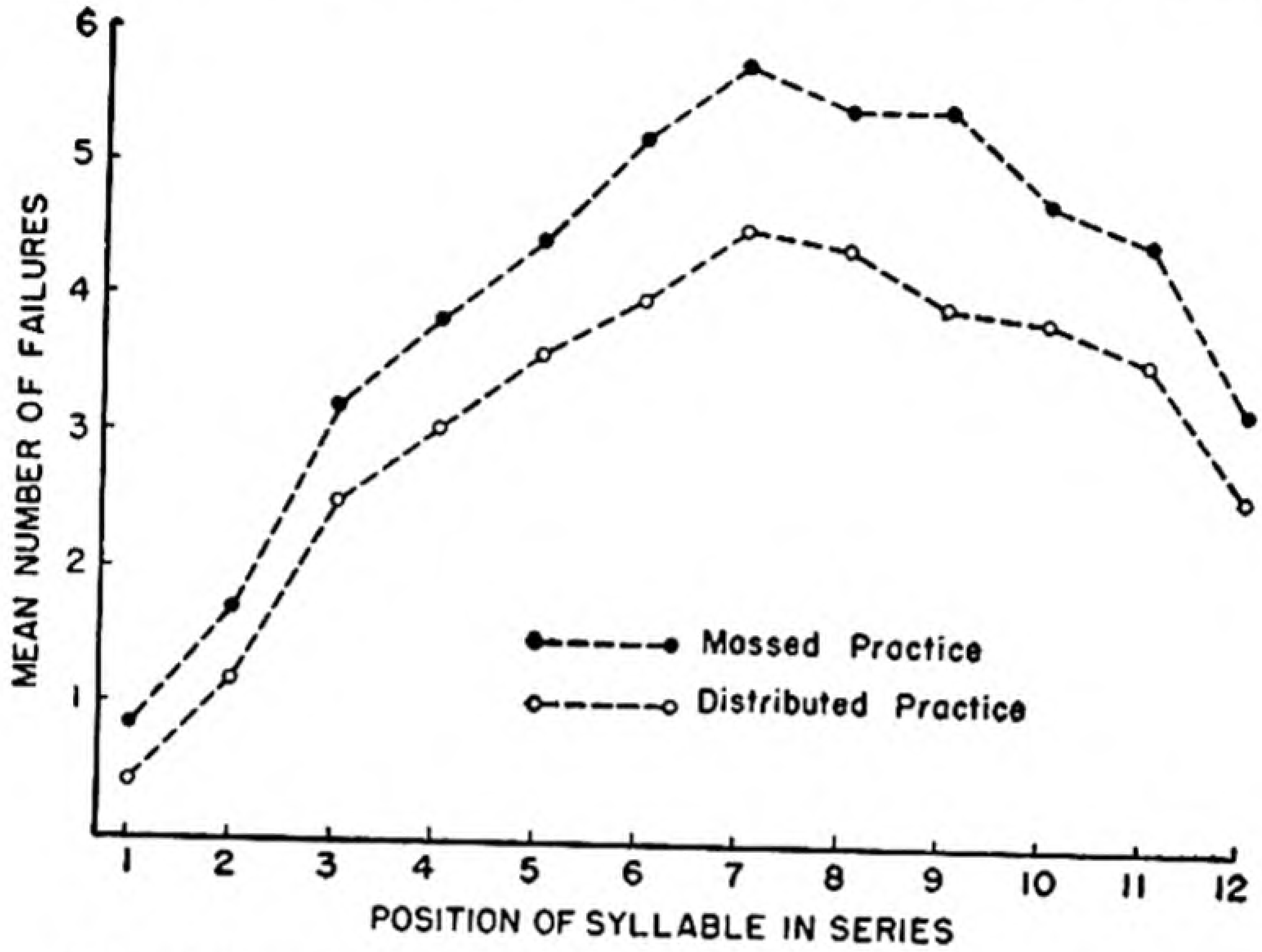


FIG. 88. Serial Position Curves Obtained Under Conditions of Massed and Distributed Practice. The serial position curve is flatter when learning is distributed than when it is massed. (From C. I. Hovland, Experimental studies in rote learning theory: I. Reminiscence following learning by massed and distributed practice, *J. Exper. Psychol.*, 1938, 22:209, by permission of the journal and the American Psychological Association.)

massed practice may allow both types of learning to develop under optimal conditions.

The beneficial effects of distribution of practice are probably due to the fact that interference among the items of a learning task is dissipated during the rest intervals. "Wrong" associations are frequently weaker than "right" ones and are readily forgotten during the rest period. For example, remote forward and backward associations may be weakened during a rest interval to a greater extent than the correct, immediate associations. Fig. 88 shows that such appears, indeed, to be the case: with distributed practice, the

serial position curve is considerably flatter than with massed practice. Not only is the series as a whole learned faster with distribution, but the items in the middle positions are the ones which are primarily responsible for the improved performance. If the interpretation of serial-position effects suggested above is correct, distributed practice does result in quicker forgetting of remote (wrong) associations and relative strengthening of immediate (correct) associations. The concept of differential forgetting allows a common interpretation of a large number of experiments showing the positive effects of distribution.

Distribution of practice may owe part of its effect to improvement in the subject's motivation. Rest periods prevent the learning task from becoming too tedious and irritating. It is doubtful, however, that recovery from fatigue accounts to any appreciable extent for the facilitating effects of distribution. Even with practice periods so short as to exclude the possibility of fatigue, distribution is still superior to massed practice.

Individual Differences Among Learners

In all spheres of behavior, individual differences among organisms constitute an important fact. This is certainly true of learning in general, and of verbal learning in particular. As experimenters, we may choose one of two alternatives in dealing with the problem of individual differences. We may try to reduce the influence of individual differences to a minimum by choosing random samples of subjects for each experimental condition and hoping that in the long run the effects of such differences will cancel out. A refinement consists of the use of matched groups of subjects, making sure that individual differences will affect different experimental conditions to fairly equal extents. In most practical experimental situations, we follow this procedure and try to minimize and randomize individual differences.

The second alternative calls for the treatment of individual differences as a systematic variable. Instead of randomizing and minimizing individual differences, we can vary them systematically, maximize their effects, as it were, and gauge their influence on observed behavior. In the field of human learning, three sources of individual differences have been systematically explored in this

fashion: age, sex, and intelligence. We shall briefly summarize the results of this work.

Learning as a Function of Age. There is good evidence that a certain amount of learning takes place before birth. Certainly, learning begins on a large scale immediately after birth and continues indefinitely from then on. To what extent does the ability to learn vary as a function of age? Clearly, learning ability depends

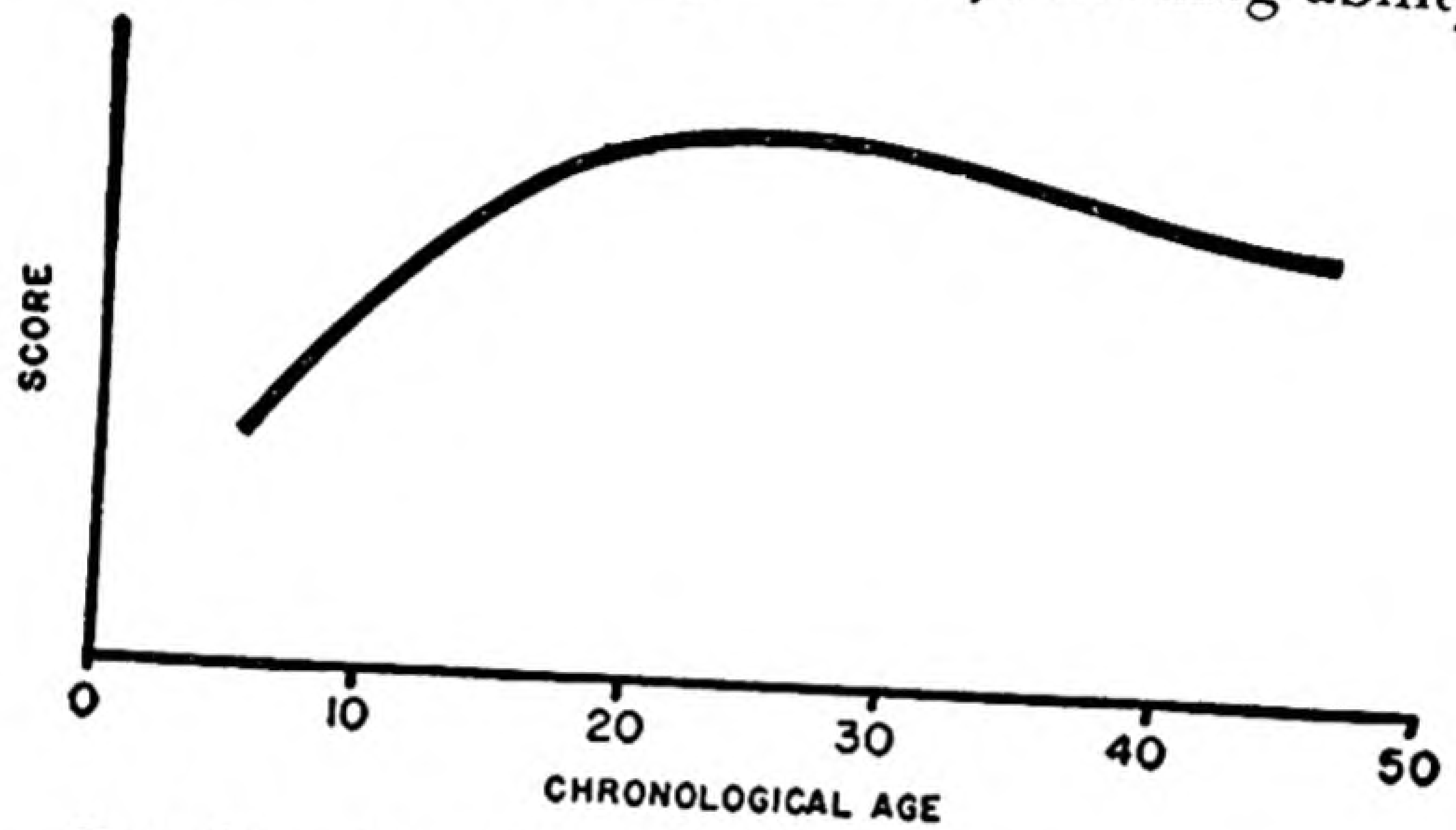


FIG. 89. Learning Ability as a Function of Age. This hypothetical curve, based on a comprehensive study of learning performance over a wide range is believed to represent the general relation between chronological age and the ability to learn. (From J. A. McGeoch, *The psychology of human learning*, 1942, p. 226, by permission of Longmans, Green & Co., Inc. After E. L. Thorndike, *Adult learning*, The Macmillan Company, 1928, p. 127, by permission of Professor Thorndike.)

on the development of certain capacities in the organism, both sensory and motor. Obviously, there can be no verbal learning before language development has reached a minimum level, although such development is itself largely a question of learning. As the organism grows and develops its capacities, learning ability increases. Systematic studies have shown that appreciable increases in learning ability do characterize the years up to late adolescence (about 18).

Does ability to learn, having reached a peak, decline? The evidence on this point is inconclusive and difficult to interpret. After maximal ability is reached, measurable decline is very slow to set

in and proceeds at a very slow rate. This finding is especially true for learning which involves the mastery of relationships and ideas. The main difficulty is that it is almost impossible to isolate experimentally "pure learning ability" in the sense of "neural plasticity." Older subjects may have learned how to learn, may be more skillful in their attacks on problems. They also have a larger repertory of information and responses which may be applied to a new problem. On the other hand, older subjects may be "out of practice." After leaving school, many of them have little opportunity to engage in formal learning activities. Thus, "pure learning ability" or "neural plasticity" are, for all practical intents and purposes, out of reach of experimental inquiry, at least with human subjects. As far as observable level of performance is concerned, there does seem to be a gradual slow decline with age, as schematically represented in Fig. 89. This graph summarizes the results of a comprehensive study of learning performance over a wide age range. Even this general trend, though supported by extensive experimental evidence, must be interpreted with caution. The task of obtaining a truly representative sample of learners of all age groups is a formidable one. Any valid generalization about learning as a function of age must be based on such a sample.

Learning as a Function of Sex. Are men or women better learners? The most valid general answer is: neither. From a large number of experimental comparisons of men and women learners, no consistent difference in learning ability between the sexes has emerged. Whatever reliable differences have been found can be ascribed to differences in interest and training. Men may be more efficient in learning materials dealing with mechanical problems; women may show greater proficiency in learning items relating to home life. Such results are, of course, due to differences in interest and training associated with sex differences. They do not reflect any inherent differences between the sexes in learning ability. To the extent that such differences in interest and training disappear or change, differences in learning ability may be expected to disappear or change concomitantly. In choosing subjects for experiments on verbal learning, it is important to ask whether or not sex differences in learning ability are likely. With nonsense materials, for example, there is no reason to expect other than random differences between

men and women learners, and it is not important that the number of men and women learners be equated under various experimental conditions. If the material is meaningful and likely to have differential appeal to men and women, sex differences need to be carefully controlled and evaluated.

Learning as a Function of Intelligence. Whenever learning ability is correlated with scores on intelligence tests, a positive relation is found. The correlation is especially high if the learning task includes abstract ideational material. We should expect such a relationship *a priori* since the score on an intelligence test is certainly not independent of the subject's previous learning ability. Those who have learned well in the past score high on intelligence tests and are also successful in handling new learning situations. One of the things which an intelligence test measures *is ability to learn*, and failure to obtain a positive relationship between intelligence-test score and learning performance would throw doubt on the validity of the test.

The practical implications for the experimenter are clear. If different learning materials or conditions of practice are to be compared, subjects must be reasonably well equated in intelligence. If it is impossible to match subjects in different experimental groups for intelligence, random samples presumed to have similar distributions of intelligence should be used. If the experimental samples are all drawn at random from a college population, systematic differences in intelligence among such samples are unlikely, especially since the population itself is characterized by a narrow range of intelligence scores.

Special Problems of Control in Learning Experiments

Certain problems of control are common to most experiments on learning. We shall list here some of the main problems of experimental control, of which we must be continuously aware.

Standardization of Instructions. We have emphasized the importance of set as a determinant of learning performance. The experimenter usually controls and manipulates set by means of instructions to the subject. Sets are highly flexible and sensitive to even slight variations in instructions. It is, therefore, important to

standardize the instructions for any given learning experiment and to keep them constant from subject to subject. Reading a standard set of instructions verbatim is the safest procedure. Never give the subject more information than is contained in the standardized instructions.

Rate and Duration of Presentation of Materials. The rate at which materials are presented to the subject and the duration for which each item is exposed should be carefully controlled. These temporal conditions should be held strictly constant for a given experiment, unless variation in rate or duration is itself an experimental variable. Speed of learning is appreciably influenced by both these factors.

The Experience of the Learner. The amount of experience which a subject has had with tasks *of the same type* as the *one on which he is tested* is a critical variable. One learns how to learn, how to attack a certain kind of problem. This is especially true for an unusual problem, such as memorizing nonsense syllables. In comparing the effects of experimental conditions, we must be sure that learners with reasonably equal amounts of experience are tested under each of the conditions. A balanced design such as the one described in Experiment XX may help to control the factor of experience.

Environmental Conditions. It is important to carry out a given experiment under constant environmental conditions (unless variations in environment are an experimental variable). In practice, this means that the whole experiment should be conducted in the same place. If an experiment comprises several sessions, it is preferable to conduct each session at the same time of day. The amount of distraction to which the learner is exposed must also be controlled. Ideally, a verbal learning experiment should be conducted in a well-lighted, quiet room. If a certain amount of noise is inevitable, an attempt should be made to keep it as constant as possible throughout the experiment.

Use of Different Experimenters. For practical reasons, it is often necessary to assign parts of the same experiment to different experimenters. Such a situation frequently arises in laboratory classes. The use of different experimenters is not likely to introduce any serious amount of error *provided all the experimenters are*

equally well trained and all follow a strictly standardized procedure. It is advisable to have each experimenter run part of the subjects under each of the experimental conditions.

Experiment XX
Serial Position Effects Under Massed and Distributed Practice

Purpose. The purpose of this experiment is to demonstrate several principles of serial learning: (1) the serial-position curve, (2) the difference in rate of learning under massed and distributed practice, and (3) the difference in serial-position effects under massed and distributed practice.

Materials. Lists of three-letter nonsense syllables. For this experiment, a reasonably short list, say, twelve syllables, will suffice. For rules to be followed in the construction of a list of nonsense syllables, see pp. 347 f.

Apparatus. The standard exposure device for verbal materials is the memory drum. This device consists of a drum mounted behind a screen. A strip of paper on which the syllables are printed is fastened around the drum. The drum is driven by a constant-speed motor and exposes each syllable for a short period of time (usually 2 seconds) through a window in the screen. A typical memory drum is shown in Fig. 90. If a memory drum is not available, the syllables may be printed on cards and presented manually. Since rate and length of exposure are of great importance, manual presentation must be carefully timed.

General Procedure. The method of anticipation should be used. The subject is seated in front of the drum, and the following instructions are read to him:

"This is a learning experiment. A series of nonsense syllables will be shown to you through the window in this screen (experimenter points to memory drum). A nonsense syllable is a combination of three letters without any meaning. I want you to look at these syllables carefully. After the list has been presented once, the syllables will be shown again, but this time it will be your task to anticipate each syllable, that is, to spell it out before it appears in the window. When the symbol marking the beginning of the list appears,[5] you are to spell out the first syllable. When the first syllable appears, you are to spell out the second, and so on through the entire list. Remember, your task is to spell out each syllable *before* it appears in the window. We shall con-

[5] A simple geometric design such as a square or triangle may be used as a cue symbol to mark the beginning of the list.

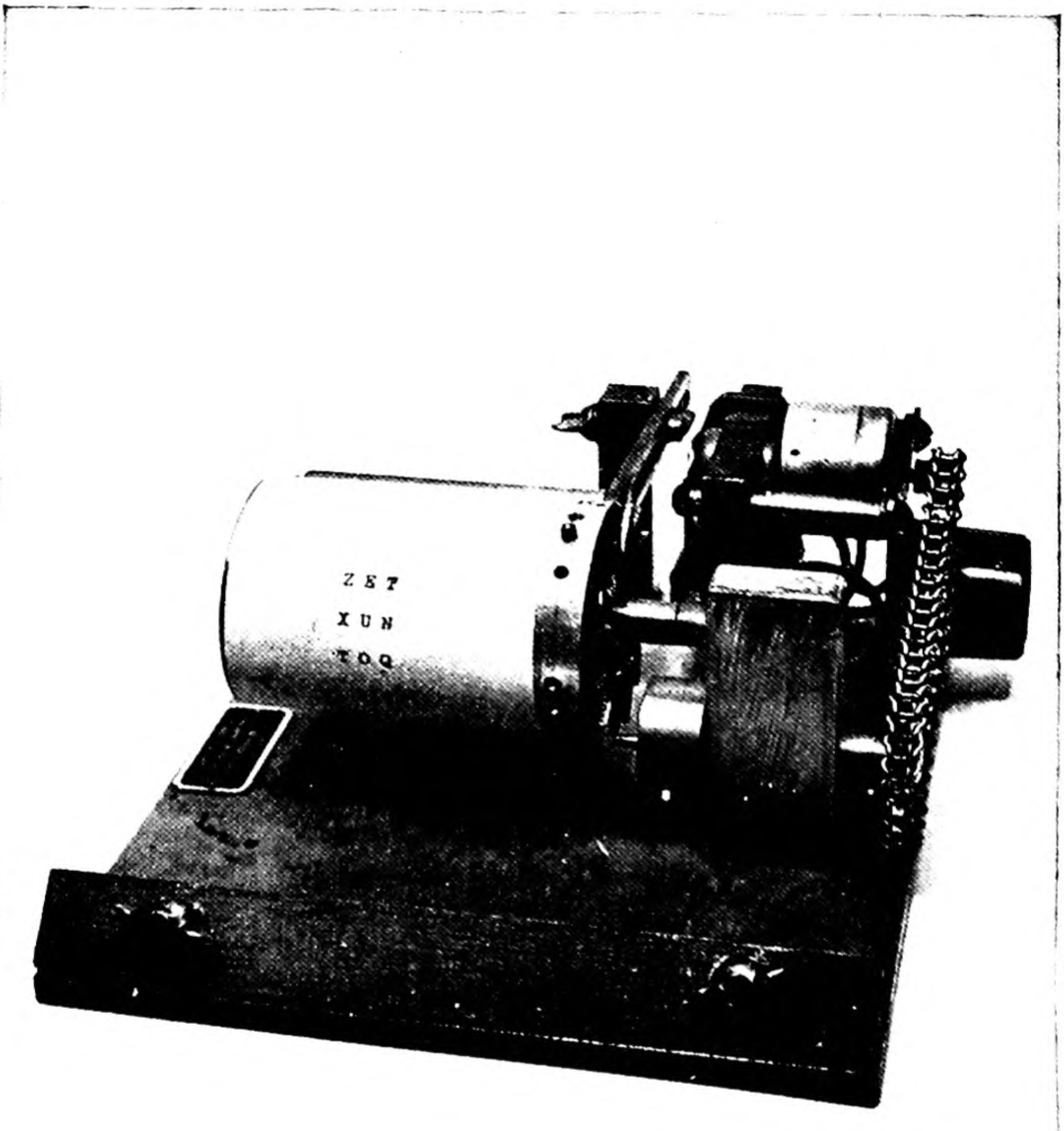
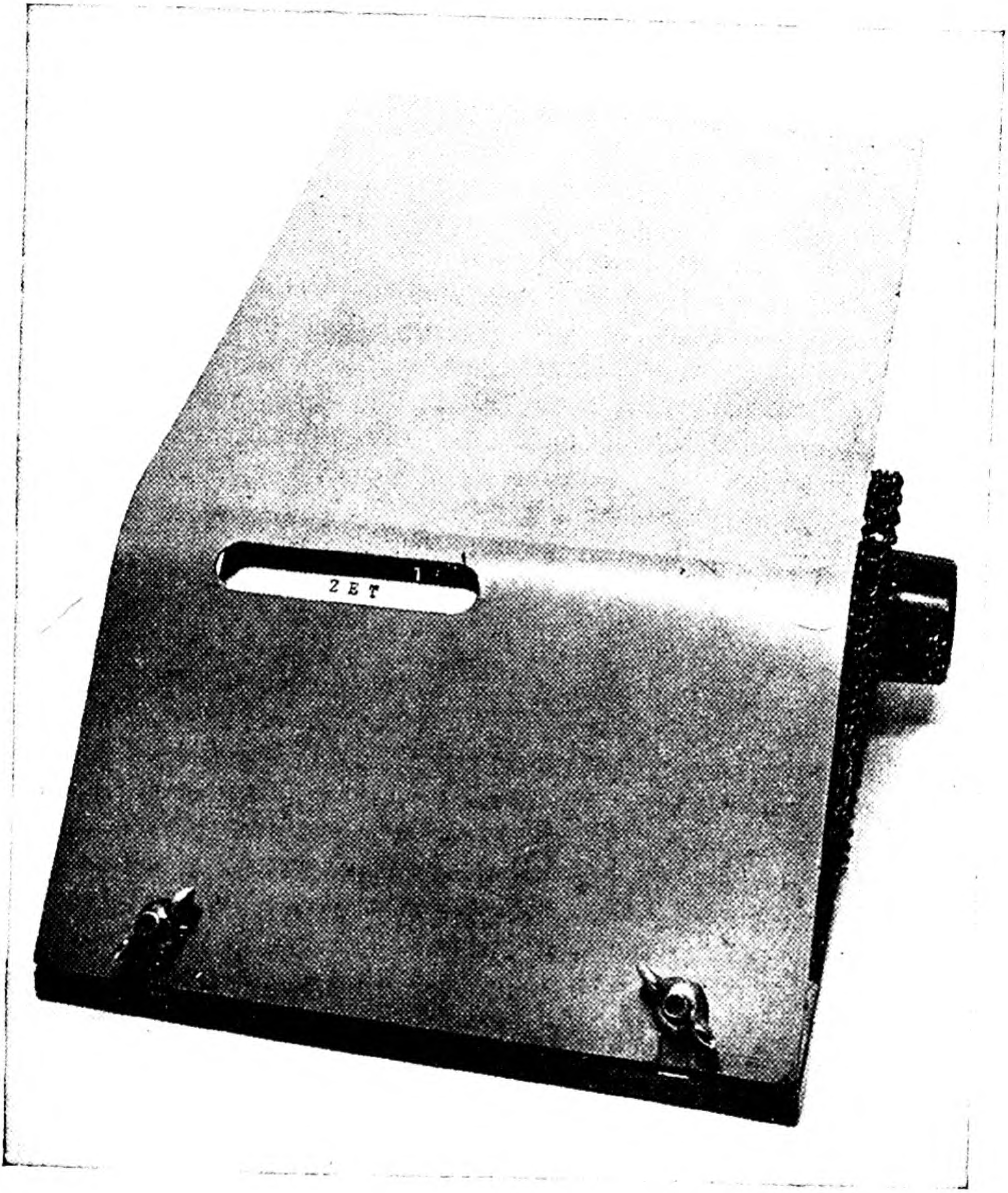


FIG. 90. A Typical Memory Drum. On the left is a view of the device with the screen removed. On the right, the drum is shown ready for use in an experiment.


ZET

tinue this procedure until you have spelled out each syllable correctly on the same trial."

After the instructions have been read, the subject should be given an opportunity to ask questions. It is essential that the procedure be quite clear to the subject before the experiment begins. When the procedure has been fully clarified, the memory drum is started. The rate at which the syllables appear depends on their spacing on the drum. Presentation at 2-second intervals is customary.

Presentation of the list is continued until the criterion is reached. As stated in the instructions, we are suggesting the criterion of complete mastery. The experimenter may, of course, choose a different criterion (a less exacting one like seven correct anticipations or a more exacting one like three errorless trials) and modify the instructions accordingly.

Massed Practice. Under conditions of massed practice, the presentations of the list follow each other in close succession. An interval of 6 seconds between trials would constitute massed practice.

Distributed Practice. For distribution of practice, rest intervals are introduced between successive presentations of the list. For this experiment, we suggest a rest interval of 2 minutes. It is important to control the subject's behavior during the rest intervals. If left to himself, the subject may use the rest interval to rehearse the list and, in effect, mass his practice. The best way to control the rest interval is to require the subject to perform a task which does not leave him any time to rehearse the syllables. For example, he may be asked to name a series of colors which appear in the window of the memory drum. Other tasks may be used, but the experimenter must make reasonably sure that the subject does not use the rest period for rehearsal.

Order of Conditions. One of the purposes of the experiment is to compare performance under massed and distributed practice. To obtain the necessary data, it is possible: (1) to divide the members of the laboratory into two groups, one of which works with distributed, the other with massed practice; or (2) to have each learner work under both conditions. If the second of these alternatives is chosen, it is, of course, necessary to use two different lists of approximately equal difficulty If each learner works under both conditions, it is also advisable to work out a balanced experimental design. Each of the two lists should be used equally often in massed and in distributed practice. Thus, differences in difficulty between the two lists will affect the two methods of practice to an equal extent and presumably cancel out. Furthermore, half the subjects should start with massed practice, the other half with distributed practice. This balanced sequence of conditions will allow both methods

of practice to benefit equally from experience and possibly to suffer equally from the effects of fatigue and boredom. The total experimental design is summarized in the accompanying table.

	Massed Practice		Distributed Practice	
Order of Conditions	First	Second	First	Second
List A	Group 1	Group 4	Group 3	Group 2
List B	Group 2	Group 3	Group 4	Group 1

Record of Results. Throughout the experiment, a careful record must be made of each individual attempt at anticipation. A typical record sheet will look somewhat as shown in the accompanying tabulation.

	Trial Number							
Syllable	1	2	3	.	.	.	.	.
1								
2								
.								
.								
.								

On each trial, and for each syllable, correct anticipations are indicated by a check mark, incorrect ones by an x (or any other system of notation, such as pluses or minuses, which the experimenter may prefer). In this manner, an accurate picture of the temporal trend of learning can be obtained.

Treatment of Results. We first compare the rate of learning under massed and distributed practice. We find for each subject the number of trials required to reach the criterion. The mean number of trials to criterion is then determined for the two conditions. Our expectation on the basis of previous experimental results is that the criterion is reached in fewer trials with distributed than with massed practice. To show the differences in rate of acquisition, learning curves showing number of correct anticipations for successive trials should be plotted. For learning curves showing the average results for several learners, the Vincent curve technique, described in Chapter 13, should be used.

Next, serial-position curves are plotted for learning under the two conditions of practice. Serial-position curves may be plotted in two ways, illustrated respectively in Fig. 84 and Fig. 88. In Fig. 84, average number of correct anticipations per trial is plotted against position in the series, yielding a typical bow-shaped curve. If a syllable were correctly an-

ticipated on all trials by all subjects, the value plotted against its position would be one.

To obtain these values, we add the total number of a subject's correct anticipations for a syllable and then divide by the total number of trials. The results of several subjects may then be averaged.

As Fig. 88 shows, the serial-position curve may also be plotted in terms of the number of failures rather than correct anticipations at each position in the series. In that case, the usual bow-shaped curve is inverted, with the peak at the middle of the series.

The serial-position curves for massed and distributed practice should be plotted against a common set of axes to make their immediate comparison easier. As Fig. 88 shows, we should expect the serial-position curve to be flatter under conditions of distribution than with massed practice. The significance of the difference between any pair of points may be tested by conventional statistical methods.

Experiment XXI

Speed of Learning for Different Amounts of Material

Purpose. To measure speed of learning as a function of the total amount of material learned.

Materials. Stimulus lists of different lengths. It is important that the lists differ in length but that the items be of uniform difficulty. The use of nonsense syllables, matched in association value, is, therefore, advisable. Meaningful words may also be used provided they are reasonably uniform in difficulty. Lists of 8, 12, and 16 items provide a convenient gradation of amount.

Apparatus. As in Experiment XX, the stimulus items are most conveniently presented by means of a memory drum. The individual items may again be presented at 2-second intervals.

Procedure. For a description of the procedure and instructions, refer to Experiment XX. The same procedure applies to this experiment. The method of anticipation is again chosen because it gives the most accurate picture of the course of learning.

Order of Conditions. We wish to compare speed of learning for different amounts of material (8, 12, and 16 items). It is again possible to have each learner learn all three lists or to divide the subjects into three groups (selected at random), each of which learns a list of different length. The latter procedure will, of course, save much time and effort.

If each learner learns all three lists, a counterbalanced design similar to the one described in Experiment XX should be used. Different learners

should practice the three lists in different orders. With three lists, six different orders of practice are possible. If enough subjects are available, an equal number of subjects should practice the lists in each of the six orders.

Records of Results. The record sheets described above should again be used.

Treatment of Results. State for each list the average number of trials required to reach criterion. We should expect the amount of time to criterion to go up as length of list increases. The main question which we wish to ask of the data is whether the average difficulty of an item increases as the total amount of material is increased. We may define difficulty in terms of *time-per-item*. Since each item is presented for a constant period of time, we can easily compute the total time required to reach criterion. Dividing the total amount of time by the number of words or syllables yields the average time-per-item for a given list. The significance of the differences between average times-per-item can be tested. The influence of amount of material can also be shown by plotting the learning curves for lists of different lengths. It can then be seen whether a practice trial at any given stage of learning does or does not result in an equal amount of improvement for different lengths of lists. If the results of several subjects are averaged, Vincent curves must be used.

APPENDIX: Rules for the Construction of Nonsense Syllables[6]

One of the most serious problems of control in experiments on verbal learning concerns the difficulty of the materials used. Nonsense syllables are so frequently used because they can be arranged in lists of fairly uniform and known difficulty. The following rules are helpful in the construction of lists of nonsense syllables.

1. The association value of the syllables must be taken into account. The norms developed by Glaze and Hull may be used here. If more than one list is used, they should be equated in average association value (unless differences in association value are themselves an experimental variable). It is advisable to construct the lists so that successive pairs of syllables have approximately the same association values.
2. No vowel should be repeated in any consecutive four syllables, except in going from the end to the beginning of the list. It is permissible, therefore, to use a vowel in one of the last three units of the list and

[6] We are indebted to Professor A. W. Melton for permission to use some of his unpublished material.

also in one of the first three units of the same list. The same vowel may not, however, be used in the first and last unit of the same list.
3. No four consecutive syllables shall have any letter (either vowel or consonant) in common. However, up to three initial and three final consonants may be the same in any given list, especially if the list contains more than twelve syllables. For shorter lists, neither initial nor final consonants need be repeated.
4. No two syllables in the same list should have two letters in common, i.e., neither one vowel and one consonant, nor two consonants.
5. Alphabetical progressions of initial and final consonants should be avoided.

Careful observance of these rules will help to make the nonsense syllable a useful type of item for experiments on verbal learning, especially where materials of uniform difficulty are desired.

General References

Hunter, W. S., Learning: IV. Experimental studies of learning. In C. Murchison (ed.), *A handbook of general experimental psychology*, Worcester: Clark University Press, 1934.

Katona, G., *Organizing and memorizing: studies in the psychology of learning and teaching*, New York: Columbia University Press, 1940.

McGeoch, J. A., *The psychology of human learning*, New York: Longmans, Green and Co., 1942.

Melton, A. W., Article on learning. In W. S. Monroe (ed.), *Encyclopedia of educational research*, New York: The Macmillan Company, 1941.

Woodworth, R. S., *Experimental psychology*, New York: Henry Holt and Co., 1938, Chaps. 1, 7, and 9.

Methods of Practice

See General References

Performance as a Function of What Is Learned

Meaning

Carlson, H. B., and Carr, H. A., Rote and logical retention memory, *J. Exper. Psychol.*, 1940, *26*:199-210.

Glaze, J. A., The association value of non-sense syllables, *J. Genet. Psychol.*, 1928, *35*:255-267.

Hull, C. L., The meaningfulness of 320 selected nonsense syllables, *Amer. J. Psychol.*, 1933, *45*:730-734.

McGeoch, J. A., The influence of associative value upon the difficulty of nonsense-syllable lists, *J. Genet. Psychol.*, 1930, *37*:421-426.

Reed, H. B., Meaning as a factor in learning, *J. Educ. Psychol.*, 1938, *29*:419-430.

Affectivity

Carter, H. D., Jones, H. E., and Shock, N. W., An experimental study of affective factors in learning, *J. Educ. Psychol.*, 1934, *25*:203-215.

Cason, H., The learning and retention of pleasant and unpleasant activities, *Arch. Psychol.*, 1932, No. 134.

Levine, J. M., and Murphy, G., The learning and forgetting of controversial material, *J. Abn. Soc. Psychol.*, 1943, *38*:507-517.

Meltzer, H., The present status of experimental studies on the relationship of feeling to memory, *Psychol. Rev.*, 1930, *37*:124-139.

Amounts Learned

Lyon, D. O., *Memory and the learning process*, Baltimore: Warwick and York, 1917.

Robinson, E. S., and Heron, W. T., Results of variations in length of memorized material, *J. Exper. Psychol.*, 1922, *5*:428-448.

Scott, T. C., and Henninger, L. L., The relation between length and difficulty in motor learning; a comparison with verbal learning, *J. Exper. Psychol.*, 1933, *16*:657-678.

Serial Position

Hovland, C. I., Experimental studies in rote-learning theory: II. Reminiscence with varying speeds of syllable presentation, *J. Exper. Psychol.*, 1938, *22*:338-353.

Hovland, C. I., Experimental studies in rote-learning theory: III. Distribution of practice with varying speeds of syllable presentation, *J. Exper. Psychol.*, 1938, *23*:172-190.

Hull, C. L., The conflicting psychologies of learning—a way out, *Psychol. Rev.*, 1935, *42*:491-516.

Lepley, W. M., Serial reactions considered as conditioned reactions, *Psychol. Monogr.*, 1934, *46*, No. 205.

McGeoch, J. A., The direction and extent of intra-serial associations at recall, *Amer. J. Psychol.*, 1936, *48*:221-245.

Robinson, E. S., and Brown, M. A., Effect of serial position upon memorization, *Amer. J. Psychol.*, 1926, *37*:538-552.

Ward, L. B., Reminiscence and rote learning, *Psychol. Monogr.*, 1937, *49*, No. 220.

Performance as a Function of How Learning Proceeds

Set and Motivation

Gates, A. I., Recitation as a factor in memorizing, *Arch. Psychol.*, 1917, No. 40.

Gibson, J. J., A critical review of the concept of set in contemporary experimental psychology, *Psychol. Bull.*, 1941, *38*:781-817.

Gilliland, A. R., The effect of practice with and without knowledge of results in grading handwriting, *J. Educ. Psychol.*, 1925, *16*:532-536.

Jenkins, J. G., Instruction as a factor in "incidental" learning, *Amer. J. Psychol.*, 1933, *45*:471-477.

Shellow, S. M., Individual differences in incidental memory, *Arch. Psychol.*, 1923, No. 64.

Thorndike, E. L., *The psychology of wants, interests, and attitudes*, New York: D. Appleton-Century Co., 1935.

Distribution of Work

WHOLE VS. PART LEARNING

Cook, T. W., Whole versus part learning the spider maze, *J. Exper. Psychol.*, 1937, *20*:477-494.

Crafts, L. W., Whole and part methods with visual spatial material, *Amer. J. Psychol.*, 1932, *44*:526-534.

McGeoch, G. O., Whole-part problem, *Psychol. Bull.*, 1931, *28*:713-739.

Pechstein, L. A., The whole vs. part methods in learning: comparison and summary, *Stud. Educ. (Yearb. Nat. Soc. Coll. Teach. Educ.)*, 1926, *15*:181-186.

DISTRIBUTED VS. MASSED PRACTICE

Cook, T. W., Massed and distributed practice in puzzle solving, *Psychol. Rev.*, 1934, *41*:330-355.

Hovland, C. I., Experimental studies in rote-learning theory: I. Reminiscence following learning by massed and by distributed practice, *J. Exper. Psychol.*, 1938, *22*:201-224.

Hovland, C. I., Experimental studies in rote-learning: III. Distribution of practice with varying speeds of syllable presentation, *J. Exper. Psychol.*, 1938, *23*:172-190.

Hovland, C. I., Experimental studies in rote-learning theory: V. Comparison of distribution of practice in serial and paired-associate learning, *J. Exper. Psychol.*, 1939, *25*:622-633.

Hovland, C. I., Experimental studies in rote-learning theory: VII. Distri-

bution of practice with varying lengths of list. *J. Exper. Psychol.*, 1940, *27*:271-284.

Lorge, I., Influence of regularly interpolated time intervals upon subsequent learning, *Teach. Coll. Contr. Educ.*, 1930, No. 438.

Learning as a Function of Individual Differences

Age

Bird, C., *Social psychology*, New York: D. Appleton-Century Co., 1940, Chap. 12.

Pyle, W. H., *Nature and development of learning capacity*, Baltimore: Warwick and York, 1925.

Thorndike, E. L., *et al.*, *Adult learning*, New York: The Macmillan Company, 1928.

Sex

Boynton, P. L., Are girls superior to boys in visual memory? *J. Soc. Psychol.*, 1931, *2*:496-500.

Conrad, H. S., and Jones, H. E., Psychological studies of motion pictures: V. Adolescent and adult sex differences in immediate and delayed recall, *J. Soc. Psychol.*, 1931, *2*:433-457.

Pyle, W. H., The relation of sex differences to the kind of material used, *J. Educ. Psychol.*, 1925, *16*:261-264.

Intelligence

De Weerdt, E. H., A study of the improvability of fifth grade school children in certain mental functions, *J. Educ. Psychol.*, 1927, *18*:547-557.

Garrett, H. E., The relation of tests of memory and learning to each other and to general intelligence in a highly selected adult group, *J. Educ. Psychol.*, 1928, *19*:601-613.

Wilson, F. T., Learning of bright and dull children, *Teach. Coll. Contr. Educ.*, 1928, No. 292.

»»» 16 «««

RETENTION AND FORGETTING — I

THE problems of learning and retention are treated under separate headings only for convenience of experimental analysis and measurement. The processes and conditions of learning and retention are continuous: one always implies the other. Learning is defined and measured as progressive improvement in performance as a result of practice. No such improvement would be possible if practice did not result in cumulative retention. Similarly, it would be senseless to talk about retention except in reference to a level of performance achieved during a period of learning. The practical distinction between learning and retention stems from the difference in the times at which performance is measured. Measures obtained during a period of acquisition—i.e., while the subject is working to reach a criterion—are *learning* scores. Measures obtained at varying time intervals after the end of practice are *retention* scores. Thus measures of learning describe the rate at which associations are formed and responses built up to adequate strength; measures of retention show how lasting the associations are, and what the changes in response strength are after a period of disuse.

The Measurement of Retention

There is no one pure measure of retention. There are several experimental operations for gauging the degree of retention, each of them valid in its own right, showing the availability of certain types of response. Measures of retention do, however, vary in sensitivity: some operations are more suitable than others for revealing small differences in degree of retention. The most common procedures for measuring retention are recall, recognition, relearning, reconstruction, and speed of response.

Recall. This procedure is best described as *active* recall. It is

the subject's task actively to reproduce the correct responses which he has acquired during the period of learning. There are several variations of active recall.

In tests by the *method of retained members,* the subject is simply required to reproduce as many items as he can remember. He may or may not have to do so in a definite order. Of course, the latter task is the more difficult one and usually requires a longer period of practice. The retention score is simply the number of items correctly reproduced. Accurate scoring depends on the extent to which the learning materials can be broken down into discrete items. Lists of nonsense syllables, digits, and discrete meaningful words offer no difficulty. On the other hand, a connected passage of prose or poetry does not provide equally clear-cut units. One cannot simply use the number of words correctly recalled, because many of them (articles, adjectives, pronouns, prepositions, etc.) frequently recur in the same passage. Moreover, such a scoring procedure would fail to show to what degree the learner has mastered the content of the passage. For these reasons, a connected passage is usually divided into "ideas" or "thought units" estimated to be of equal difficulty, and the final score is determined by the number of such thought units correctly reproduced.

The *memory-span method* is closely related to the method of retained members. The subject is presented (visually or orally) with a list of items, such as digits, letters, or words, and is required to reproduce them immediately after presentation. The number of items is successively increased until the subject fails in his attempt at reproduction. The longest series which he can reproduce without error defines his memory span. A more reliable measure of the span can be obtained by application of the psychophysical method of constant stimuli (see Chapter 2). By this procedure, a graded series of lists is repeatedly presented to the subject. The percent of the trials on which each length of list was correctly reproduced is tabulated. The memory span is defined as that length of list which can be reproduced correctly on 50 percent of the trials. Obtained in this manner, the span is closely akin to a psychophysical threshold. It is a threshold defining the limit of immediate retention. It goes without saying that the magnitude of the span varies widely with

the nature of the stimuli, the speed of presentation, the subject's set and other experimental conditions.

The method of anticipation, previously discussed in detail (see pp. 313 f.), is another procedure in which retention is measured by active recall. It is peculiarly characterized by the fact that each item must be reproduced in response to a specific stimulus: the item immediately preceding it in the series. The retention score is given by the number of correct anticipations.

The method of paired associates (see pp. 314 f.) also requires active recall by the subject. Here one member of a pair becomes the specific stimulus for the reproduction of the second member. The number of correctly completed pairs is the retention score.

Tests of active recall constitute a very exacting method of measuring retention. Only those responses which have acquired sufficient strength to be available for active reproduction can contribute to the subject's score. Weaker associations have little chance to manifest themselves in active recall, although they may aid other types of retention performance.

The minimum strength which a response must have to be actively reproduced is designated as the *threshold of recall*. Like all thresholds, the threshold of recall is not a fixed stable value but fluctuates in time. As a result, an item which is not available for active reproduction at one moment may emerge above the threshold of recall a few moments later. Such *oscillations at the threshold of recall* frequently occur in retention tests for verbal materials. On successive retention tests, new items not previously recalled may appear. New items may emerge above the threshold of recall even if the total amount of retention is going down.

Active recall, then, does not usually yield a maximum measure of retention. Not only must responses have acquired considerable strength in order to be actively reproduced, but what is available for recall is subject to appreciable temporal fluctuations.

Recognition. In a recognition test, the subject is confronted with a series containing both correct (previously learned) and incorrect (new) items, arranged in a random order. His task is to pick out the items which he recognizes as correct. Such a test inevitably invites guessing on the part of the subject. Some individuals are more inclined to guess than others, and a correction for guessing must be introduced in order to render the scores of different sub-

jects comparable. The scores are adequately adjusted by subtracting the percent of items incorrectly recognized from the percent correctly recognized. Suppose, for example, that the original learning task consists of twenty nonsense syllables. These twenty syllables, mixed with forty new ones, appear on the recognition test. Subject *A* recognizes ten syllables correctly but also makes four wrong identifications. His score is $10/20 - 4/40 = 50\% - 10\% = 40\%$. Subject *B* also recognizes ten syllables correctly but in addition makes twenty incorrect recognitions. His score is $10/20 - 20/40 = 50\% - 50\% = 0\%$. Thus, with proper adjustments for incorrect responses, the same number of correct "recognitions" may result in widely differing retention scores. False recognitions are not necessarily the result of random guessing but may be due simply to poor learning so that the subject is unable to discriminate between correct and incorrect items. A penalty for wrong responses is still in order if a fair estimate of retention is to be made.

There is no rule specifying the ratio of correct to incorrect items in a recognition test. An equal number of correct and incorrect stimuli, mixed at random, is frequently used. Other ratios are, however, equally permissible. The larger the number of incorrect items, the more difficult it becomes to hit on a correct response through lucky guessing. A one-to-one ratio of correct and incorrect items would thus appear to be a minimum, but larger proportions of wrong items may be used with profit.

A recognition test may or may not be a successful test of retention. If there is little or no similarity between the old and the new items, the test will be much too insensitive to measure retention. An extreme example will illustrate the point. If our subject learns a list of nonsense syllables, and we test his retention for them by mixing these syllables with an equal number of words, he will certainly identify all the syllables and reject all the words. Obviously, retention for nonsense syllables can be tested best by mixing them with other nonsense syllables, retention for adjectives by mixing them with other adjectives, and so on. The sensitivity of the recognition test will depend upon the degree of similarity between the old and new items.

Recognition performance has been found generally superior to active recall when retention scores were compared after constant amounts of practice. The ability to recognize an item seems to

depend on a lesser degree of learning than the ability to reproduce an item actively. This result is not surprising if we consider the differences between the two test situations. All the correct items appear on the recognition test. Those items which have been poorly learned may be strengthened by the very fact of their reappearance as test stimuli. Many weakly learned items, though not mastered well enough for active recall, still have greater strength than wrong items and are, therefore, correctly recognized when contrasted with entirely new items. In active recall, the learner has no opportunity to be exposed once more to the correct items, nor can he benefit from the difference between weakly learned correct items and even weaker wrong items. Much of what the learner has acquired during practice cannot manifest itself in recall because it falls short of the minimum degree of mastery required for active recall. The method of recognition is more sensitive because weaker associations have a better opportunity to contribute to the subject's performance.

Relearning. Retention can be measured by the speed with which a subject can relearn a task to criterion some time after the cessation of practice. The greater his retention, the faster he can reëstablish his old level of performance. If his retention were perfect, he would reach criterion on the first trial of relearning. If there were no retention, it would take him as long to relearn the task as it took him originally to learn it. Over a long time interval after the end of practice neither of these extreme cases is likely to occur. Relearning usually takes less time than original learning, and the amount of time or number of trials *saved* is an index of the degree of retention. Hence, this method of measuring retention is also known as the *method of saving.*

A simple example will illustrate the use of this method. A subject learns a list of words in twenty trials to the criterion of one perfect repetition. A week later, he relearns the same list to the same criterion, this time in twelve trials. He has saved eight trials or 40 percent (8/20). Similar computations of saving can be made for measures of time, numbers of errors, or whatever other indices of learning are employed. Percent saving is given by the following formula.

$$\frac{\text{Measure for Original Learning} - \text{Measure for Relearning}}{\text{Measure for Original Learning}} \times 100$$

Use of the method of saving is not without methodological difficulties. The subject necessarily relearns the task some time after the original learning. During this interval, he may have become a more practiced learner and he may have acquired more efficient modes of attack on the problem. Part of the saving may, therefore, be due not to retention of the original task but to greater learning ability. This objection may be met, in part at least, by comparing relearning not with original learning but with the learning of an equivalent new task. Returning to the example above, we recall that our subject learned the original list in twenty trials, and a week later, relearned the same list in twelve trials. Instead of using his original learning as a basis of comparison, we should have the subject learn a new list matched in length and in difficulty with the original list. Suppose it takes him eighteen trials to learn this new list. We would then conclude that he has saved $(18 - 12)/18 = 33.3\%$. This measure would probably be more accurate since it would take the change in general practice level into account. The validity of this procedure depends, of course, on our ability to construct a new task strictly equivalent to the original learning task. With nonsense materials, equivalent lists can usually be constructed without too much difficulty. With meaningful materials, the problem of obtaining equivalent tasks is a much more difficult one.

The special value of the method of relearning lies in the fact that it measures retention independently of the *availability* of specific responses. A subject may not be able to recall actively a single item and yet show a substantial amount of saving. Indeed, saving has been reported for relearning after more than twenty years.

Reconstruction. The method of reconstruction is related to the method of saving but puts its main emphasis on the retention of serial position or order in general. The learning task consists of a series of items arranged in a definite order. Some time after the end of practice, the subject is given the items he has learned, but they are scrambled in a random arrangement. His task is to reconstruct the original order. The retention score is based on the number of items put in their proper position in the series.[1] The method of reconstruction merges into the method of saving if the subject is

[1] A rank-order correlation coefficient may be used to quantify the correspondence between original series and reconstructed series.

allowed to continue his efforts until he has reëstablished the original order. The saving in time or trials is then computed as usual.

Reconstruction yields a very specialized measure of retention because of the exclusive emphasis put on sequence or order. In this stress on sequence, it is probably most closely allied to the method of anticipation. Like the latter (and recognition), the method of reconstruction can be used only with materials which can be readily broken down into separate units.

Speed of Response. The measures discussed so far are based primarily on the *amount* of work achieved by the subject on the test of retention, e.g., the number of items correctly recalled, recognized, and reconstructed. It is possible to gain additional information about degree of retention by observing the subject's behavior during the test, i.e., by examining the nature of the performance as well as the results of performance. The *temporal* characteristics of the responses in particular are sensitive indices of the degree of retention.

First of all, there is *latency*. How quickly upon presentation of the test stimulus is the response given? Students of animal conditioning have frequently used latency as a measure of the strength of conditioning (see p. 301). In many situations, latency is inversely related to strength: the smaller the latency, the greater the strength of the response. This measure is also applicable to verbal learning, especially to the method of paired associates. The speed with which the missing member of the pair is supplied by the subject is a measure of the *availability* of the response, and an index of the strength of retention supplementary to measures of sheer amount retained.

Another temporal index of strength of retention is the *rate* at which responses are given, i.e., the speed with which they follow each other. Again, this measure has been found most useful in animal conditioning to measure the strength of conditioning. Measurements of rate are also applicable to verbal learning, especially to active recall (method of retained members). The faster the rate at which the correct items are reproduced, the greater the availability of the responses. Thus rate, like latency, is a measure of retention which supplements measures of amount.

Temporal measures may yield significant differences even when

there are no significant differences in the amount of material retained. Thus two individuals may recall the same number of items, but one may recall them at a faster rate than another. The faster rate, then, indicates greater availability of responses and, by inference, a higher degree of retention.

The Temporal Course of Forgetting

The most outstanding characteristic of retention is its decrease in time. With a few exceptions, the amount of retention becomes less and less as the time interval between the end of practice and the

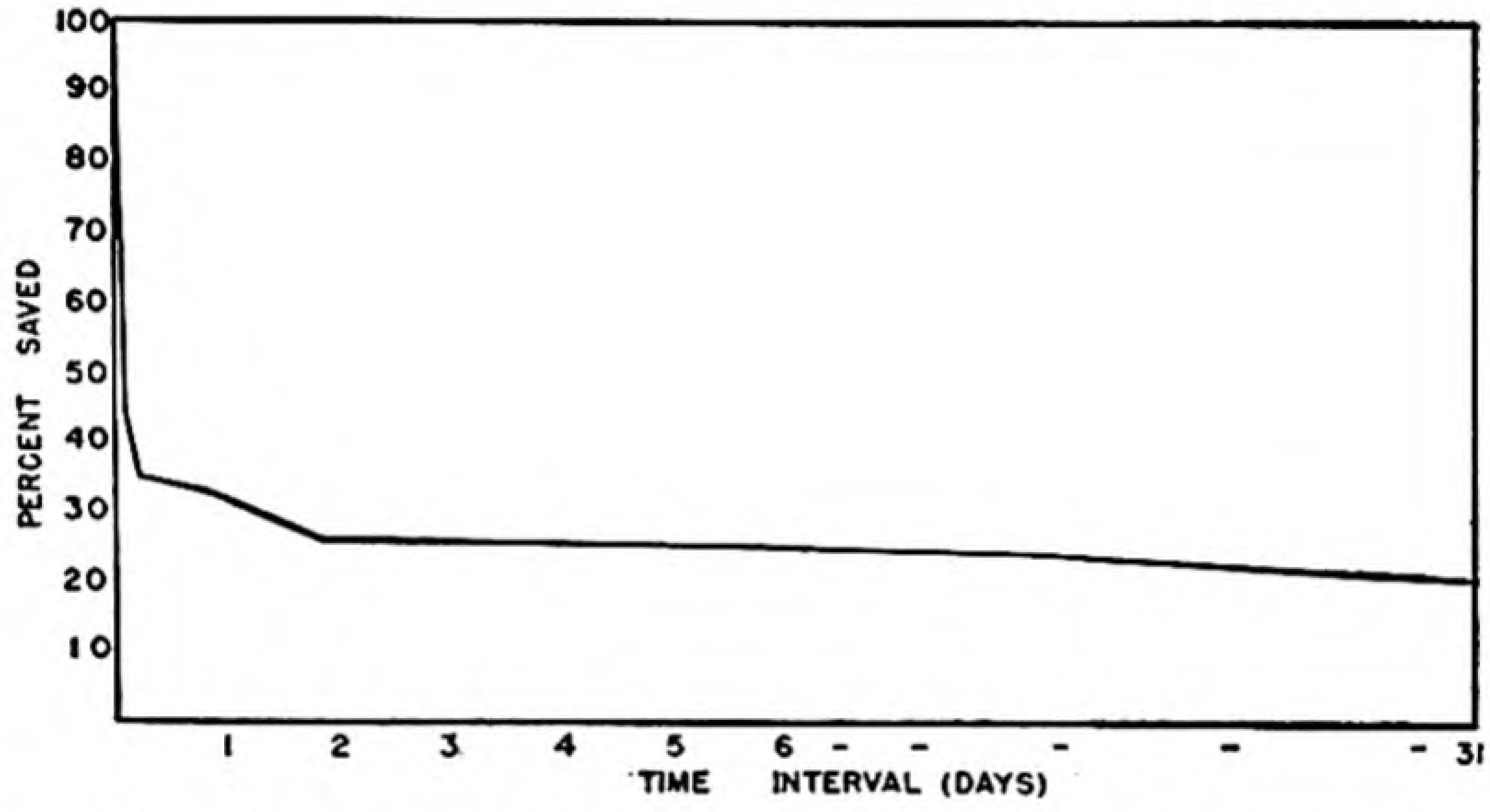


FIG. 91. The Course of Forgetting in Time. In this curve retention for nonsense syllables, as measured by percentage time saved in relearning, is plotted against time. (From data of H. Ebbinghaus, *Memory: a contribution to experimental psychology.* Trans. by H. A. Ruger, and S. S. Bussenius, New York: Columbia University, 1913, p. 76.)

memory test is lengthened. When amount retained is plotted against time, we obtain a curve of forgetting. Fig. 91 shows a forgetting curve which is a classic in the field of memory: it represents the work of the first experimenter in the field of human learning and memory, Hermann Ebbinghaus. Based on experiments performed more than sixty years ago on one subject (Ebbinghaus himself), it has stood the test of time amazingly well and has in its main features been borne out by much subsequent work done with larger samples and considerable refinements of procedure.

Ebbinghaus' curve shows retention, as measured by relearning or saving, after varying intervals of time following the end of practice. The most striking feature of the curve is its decelerated drop. Retention loss is considerable during the first few hours—after 1 hour there is a saving of only 44.2 percent; after 8 hours, the amount saved has dropped to 35 percent. After the first day, the decline is much slower and the slope of the curve is very gentle. Although there is no one, truly typical curve of retention, the decelerated decline is a feature common to curves obtained under many different experimental conditions.

The gentle slope of the later part of the curve indicates a fact which has been independently confirmed: forgetting only very, very slowly, if ever, becomes complete. As we have mentioned above, a certain amount of saving has been demonstrated in some cases even after decades. There is also "anecdotal" and clinical evidence about the revival of remote childhood experiences in old age and at moments of extreme stress. In spite of the sketchy nature of the evidence, it is probably reasonable to conclude that some effects of past learning persist indefinitely if only our measures are sensitive enough to detect them.

The Determinants of the Rate of Forgetting

We forget in time, but how much we forget and at what rate is not determined by the sheer passage of time but by a multitude of conditions which exercise their effects in time. Amount of forgetting is a joint function of a multitude of specific conditions, many of which can be varied experimentally. Proceeding, as it were, chronologically, we can localize and study determinants of retention (1) at the time of original learning, (2) in the time interval between the end of practice and memory test, and (3) at the time of the retention test.

Retention as a Function of the Conditions of Learning

The basic continuity of the processes of learning and retention would lead us to expect that the degree to which learning is carried, the nature of the materials learned, and the conditions under which learning takes place are important determinants of the course of

retention. This expectation is supported by a considerable amount of experimental evidence.

Degree of Learning. Measures of retention gauge how *lasting* the effects of learning are. That which is strongly fixated should last longer, and it has, indeed, been found experimentally that the higher the degree of original learning, the greater the retention. In some cases, this relationship approaches linearity. Over a considerable range, the increase in retention of nonsense syllables, as meas-

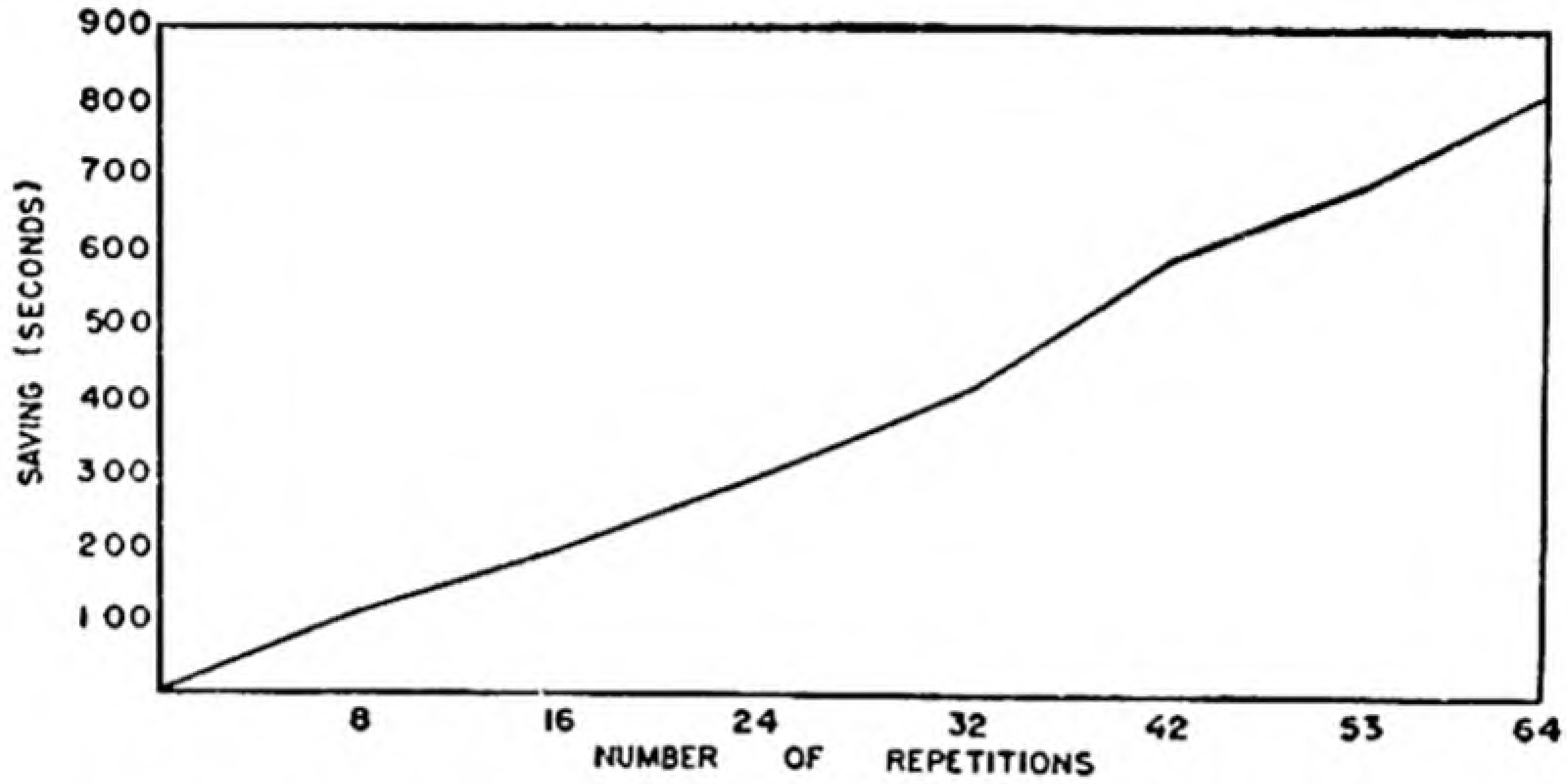


FIG. 92. Retention as a Function of the Number of Repetitions in Learning. This curve shows retention, as measured by the number of seconds saved in relearning a list of nonsense syllables, plotted against the number of repetitions during original learning. (Data from H. Ebbinghaus, *Memory: a contribution to experimental psychology*. Trans. by H. A. Ruger and S. S. Bussenius, New York: Columbia University Press, 1913, p. 56.)

ured by relearning or saving, for example, may be directly proportional to the number of times the list was originally repeated. Fig. 92 illustrates this point.

Amount of retention cannot increase indefinitely as a function of degree of learning. A point of diminishing returns is reached. If the material is poorly learned, additional repetitions will considerably increase the amount of subsequent retention. Overlearning, i.e., repetitions of the material after the criterion of complete mastery has been reached, will further increase retention. But more and more overlearning soon yields only small increments in retention. In general, then, amount of retention is positively correlated with

degree of learning, but the higher the degree of learning, the less is retention likely to benefit from further learning.

In relating retention to degree of learning, the latter may be varied in several ways. Number of practice trials provides one way for varying degree of learning: the larger the number of repetitions, the higher, presumably, the degree of learning. A constant number of repetitions, however, results in different degrees of mastery from subject to subject, and, therefore, number of repetitions may have different effects on the retention performance of different subjects. It is often desirable, therefore, to define degree of learning in terms of the criterion reached by the subjects, such as one perfect repetition, two perfect repetitions, and so on. Retention may then be measured following different criteria of mastery.

Not even a constant criterion of mastery insures a constant degree of learning. When two individuals can satisfy a criterion of, say, one perfect repetition, we can only conclude that both subjects have at the time of testing reached the minimum degree of mastery necessary to satisfy the criterion. They may still differ in degree of learning as measured by more sensitive tests. Whether or not equal test performance means equal degrees of learning depends, among other things, on the *age of the associations* involved. Suppose individual A learns a list to criterion at time t_1 and relearns it to the same criterion at time t_2. Individual B learns the same list to the same criterion but does so only once, at time t_2. At time t_2, then, both subject A and subject B show the same degree of learning as measured by their ability to reach the criterion. On a test of retention some time after t_2, however, individual A is likely to be superior to B. His learning (associations) of the list are older and "stronger." The dependence of retention on age of association is formally known as *Jost's law: Of two associations which meet the same criterion, the older one diminishes less with time*. In other words, each relearning of the material to a constant criterion increases the degree of learning, leads to longer retention. In the light of Jost's law, the ability to reach a criterion is not an unequivocal index of degree of learning. Whenever degree of learning is used as an independent variable, as in measuring its effect on retention, the effects predicted by Jost's law must be borne in mind and, if possible, controlled (e.g., by using associations of equal age).

Distribution of Practice. We have already noted (pp. 332-336) that retention is better after distributed than after massed practice. The superiority of retention after distribution follows immediately from Jost's law. Distribution of practice involves the interpolation of rest intervals during practice. Thus, when a criterion is reached by distribution of practice, the associations are older than in the case of massed practice. The factors which lead to faster learning by distributed practice (e.g., the forgetting of incorrect, interfering associations) are probably also responsible for the superior retention following this method of practice.

Amount of Material. Regardless of the specific conditions of practice, degree of retention varies with the length of the task. We saw before (see pp. 320 f.) that a long task is more difficult to learn than a short one; when the task is long, the learner has to spend a larger amount of time on each item than in the case of a short task. Once learned, however, the long task is better remembered than the short one. When two lists of different lengths are learned to the same criterion of mastery, a larger percentage of items is retained from the longer of the two lists.

The direct relationship between amount of material and the relative degree of retention may at first be surprising but is a logical consequence of the greater difficulty of the longer task. A long task requires a long period of practice, and a considerable amount of time is spent on the individual items as well as in the effort to connect and organize them. As a result, many of the items are overlearned. Having been overlearned, they have higher retention value than the items in a short list which are easier to learn and require a briefer period of practice. The dependence of retention on the length of the task thus reduces in large part to the effect of another basic variable: degree of learning.

Nature of the Learning Task. Rate of learning varies for different kinds of activities and so does rate of forgetting. Meaningful materials are better retained than nonsense items. The general content and meaning of a passage are retained better than the verbatim statements. In general, retention favors vivid and distinct experiences that are rich in associative support.

The fact that vivid and unusual experiences are favored in memory provides a valuable cue to the way in which retention depends

on the nature of the materials. A series of stimuli, all of which are homogeneous in form and quality, such as a series of nonsense syllables, is retained only with great difficulty. After the end of practice, the discrimination among these highly similar items, which was difficult enough while they were physically present, becomes more and more inaccurate. As the Gestalt psychologists put it, the traces of homogeneous or "crowded" materials cannot maintain their identity for long. Because of their similarity, there is maximum opportunity for mutual interference among the traces of the individual items. On the other hand, any stimulus or experience which stands out from its background, which is "isolated" in a series by virtue of its form and quality, is heavily favored in retention.

It is possible to demonstrate the greater retention value of "isolated" material by a simple experiment. The stimulus list consists of a series of items, most of which are of the same general form, say, nonsense syllables. Interspersed among the syllables are a few items of an entirely different nature, e.g., three-digit numbers, geometric figures, etc. After a constant amount of practice, the isolated items will invariably show a higher degree of retention. As a necessary control, the relation between the two kinds of materials should be reversed in another series: the numbers or geometric figures are "crowded" and the nonsense syllables are "isolated." If the isolated items are still favored in retention, that is obviously due to the very fact of their being isolated and not to the intrinsic ease or difficulty of nonsense syllables, geometric designs, etc. This general type of experiment has been performed repeatedly and the results have clearly established the superior retention value of isolated materials in a homogeneous series.

Motivational Determinants

In discussing the retention value of different types of materials, the learner's attitude toward the materials, his set, his interests, and emotional responses are an ever-present source of variation. We have already discussed these factors as variables in learning experiments (see pp. 326-331). They play an equally important role as determinants of the course of retention.

Set. Without a set to learn (whether induced by the experimenter or by self-instruction), little or no learning takes place (see

pp. 328 f.). Since retention varies with degree of learning, only those learners who practice under an adequate set will show good retention. A specific set to recall will aid retention further. A subject who expects a retention test after the end of practice will do better than one whose memory is tested unexpectedly.

The more fully a learner knows *how* his retention will be tested, the better will be his performance. If a subject knows in advance that his retention will be tested by a particular method, say, anticipation or recognition, he will attempt to practice the material so as to be maximally prepared for the test. If he knows that there will be an anticipation test, he will emphasize serial position during the period of practice. If he expects a recognition test, he will disregard position and concentrate on the nature of the individual items. Performance is optimal if the subject knows the nature of the retention test and prepares himself while practicing the material. If the expected test and the one actually administered differ, retention is impaired. For example, if the subject practices the material with a view to an anticipation procedure and is then confronted with a recognition test, his performance is poorer than if he had practiced specifically for a recognition test. Learning always involves selective emphasis on certain aspects of the stimulus material (such as serial order, meaning, etc.). The learner's subsequent performance depends on the extent to which the retention test calls for the same kind of response that he selectively emphasized during the period of practice.

Interests, Attitudes, and Values. Explicit instructions are by no means the sole determinants of selective retention. The subject's reaction to the material is important: the degree of his interest in it, the attitudes which it evokes, the way in which it fits into his system of values. We have already considered such factors as determiners of the rate of learning; they continue to exercise their influence as determinants of the course of retention. Although these variables have not been too well explored, certain general conclusions have already begun to emerge.

1. The greater a subject's personal interest in the material which he has learned, the better will be his retention for it over a considerable period of time. A measure of interest may be obtained by requiring the subject to rate the learning material on a scale of

interest (ranging from very interesting through indifferent to very dull), but unfortunately the validity of such ratings is not too well established. The degree of interest in the learning material may also be estimated from known characteristics of the learner, such as his occupation and his social background.

2. Attitudes of acceptance and rejection toward the content of the learning material may be significant determinants of the degree of retention. If the material learned is controversial (concerning, for example, a political or religious issue), subjects who differ radically in their attitudes toward the controversial issue are also likely to show differential retention for such material. There is some experimental evidence showing that material which is in agreement with the subject's own view is favored in retention. Similarly, it has been shown that material which is probably experienced as hostile or threatening to established values and attitudes may be retained more poorly than material which is acceptable to the subject or toward which he is indifferent.
3. The success of experiments studying the relationship between attitudes and values on the one hand and retention on the other depends on the investigator's ability to find materials which do indeed have important personal relevance for the subject. One of the most successful studies in this area was one in which the subjects were psychiatric patients, and emotionally toned learning materials were selected on the basis of these very patients' clinical records. Items which these records suggested to be acceptable to the subjects were retained significantly better than items similarly classified as indifferent or unacceptable.

The experimenter must guard against assuming lightly that his subjects will share his own judgments of pleasantness and unpleasantness, his own feelings of acceptance and rejection toward the learning materials. Experiments on the relation of attitude to retention may easily founder on the failure to select materials of true personal relevance to the subjects.

Memory for Completed and Interrupted (Successful and Unsuccessful) Tasks. The role of motivational determinants is clearly illustrated by the difference in retention for completed and interrupted tasks. In a typical experiment, subjects are given a series of tasks (usually simple manual tasks) to perform. They are per-

mitted to complete half these tasks while the other half is interrupted by the experimenter and remains uncompleted. At the end of such an experimental session, the subjects are required to recall as many of the tasks on which they had worked as they can remember. Frequently and reliably, a higher percentage of interrupted than completed tasks is recalled.[2]

The superiority of interrupted over completed tasks can be quantitatively expressed by means of the *I/C* ratio, where *I* equals number of interrupted tasks recalled and *C* equals number of completed tasks recalled. If there is no difference in the retention value of the two types of task, the ratio is equal to 1. The greater the ratio, the more are interrupted tasks favored in retention. *I/C* ratios as high as 1.9 have been reported. A valid *I/C* ratio depends, of course, on adequate control of factors other than completion and interruption of the task. A counterbalanced design (see p. 345) must be used, so that each task appears both in the completed group and the interrupted group an equal number of times. In this manner, it is impossible for the nature of the task as such (its familiarity, vividness, etc.) to affect the *I/C* ratio.

Whether or not an *I/C* ratio greater than unity is obtained depends to a considerable extent on what the fact of interruption means to the subject, how his motivation in performing the tasks is affected by the interruption. Subjects may interpret interruption as a sign of either success or failure in the task. If the interruption is made without explanation or under some vague pretext, many subjects are likely to interpret it as an indication of failure in the task. Under these conditions, the *I/C* ratio is often high. On the other hand, subjects may be led to believe that being interrupted in a task means that they have succeeded in it. In experiments in which interruption stood for success, the *I/C* ratio is often considerably diminished and may even drop below 1.

Interrupted tasks owe their retention value, at least in part, to their interpretation as success and failure by the subjects. Such a statement, however, only serves to raise a further question: How do successful and unsuccessful tasks differ in their retention value? The experimental answers to this question have often been

[2] Subjects not only retain a higher proportion of uncompleted tasks but also tend spontaneously to resume them if given an opportunity to do so.

contradictory but order emerges out of the data when the characteristics of the subjects who succeed and fail in the tasks are taken into account. Preferential retention for successes and failures is significantly correlated with personality characteristics of the subjects and with the importance which they attach to success and failure in any particular tasks. Thus, the more "proud" the individual, the more he feels his self-esteem threatened by failure, the more likely he is to remember a greater proportion of successes than failures. Similarly, the more important and significant he believes the task to be as a measure of his ability, the more likely he is to favor successes over failures in memory. The measurement of independent variables such as "importance" and "threat to self-esteem" is often difficult and involves the use of ratings and judgments, the reliability and validity of which are far from perfect. Nevertheless, even rough classifications of subjects according to these variables show important differences in the retention of completed and interrupted, successful and unsuccessful tasks. Relationships such as these underscore the importance of considering the motivation of the subject in studying the determinants of retention.

Retention as a Function of Interpolated Activity: Retroactive Inhibition

We have studied degree of retention in relation to the conditions under which the original learning of the given material takes place. We now turn to a second group of determinants—the nature of the activities which fill the interval between the end of practice and the test of recall. We must emphasize again that time in and of itself does not do anything and is not a useful independent variable in the study of retention any more than it would be a useful independent variable in the interpretation of physical and chemical processes. Time provides us with a framework of measurement within which the lawful determinants of retention and forgetting take their course. Among these determinants, the activities between learning and recall are of paramount importance.

The Concept of Retroactive Inhibition

Retention After Sleep and Waking. The importance of the activity filling the interval between learning and recall can be

dramatically illustrated by comparing the effects upon retention of periods of sleep and waking. Periods of sleep and waking, which differ so clearly in amount of activity, affect retention in strikingly different ways. A period of sleep immediately following the end of practice greatly reduces the amount of forgetting as compared with a period of wakefulness. In one experiment, for example, subjects learned a list of nonsense syllables to a criterion of perfect mastery and then were tested after periods of 1, 2, 4, and 8 hours spent either in sleep or in ordinary waking activities. At all points tested, a period of sleep resulted in superior retention. The differences are especially striking after the longer periods. There is *some* forgetting during the first few hours of sleep but thereafter the amount of retention remains almost unchanged, whereas a period of wakefulness yields a steady downward trend (forgetting curve).

An interesting parallel to these findings with human subjects has been established with animals and, indeed, with a very humble experimental subject—the cockroach. A cockroach can learn a simple discrimination response, such as turning toward a bright light. Immediately after the training period, the insect may be completely immobilized by being put into a small dark box. Following such a period of immobility, the discrimination is retained considerably better than after an equal period of physical activity. As in the case of human subjects, a period of physical inactivity protects the organism from forgetting. During such a period, few events can take place which would interfere with the results of learning.

It is true that some forgetting does take place during sleep, especially during the first few hours after the end of practice. Even sleep does not mean the absence of all activity. A certain period of time between the end of learning and the full onset of sleep cannot be controlled. A certain amount of activity continues during sleep as the presence of dreams well illustrates. Even if intervening activities were the sole causes of forgetting, retention could not be expected to be perfect after sleep. It would be rash, however, to make such a sweeping generalization about the causes of forgetting. What the superiority of performance following sleep does demonstrate is that the activity between learning and recall is a major determinant of the degree of retention.

Definition of Retroactive Inhibition. The adverse effect upon retention of an activity interpolated between learning and recall is

described as *retroactive inhibition.* The concept may best be defined by the experimental operations which are used to establish and measure it.

Suppose there are an experimental group and a control group well matched in learning ability. Both groups learn a task, *A,* to a criterion. We then treat the two groups differently during the retention period. The experimental group learns another task, *B,* which often bears a certain amount of similarity to *A* (e.g., *A* and *B* may be two different lists of nonsense syllables). While the members of the experimental group are learning *B,* the members of the control group "rest," i.e., they do not learn anything but are occupied with something which bears as little relationship to *A* as possible (e.g., if task *A* was learning a list of nonsense syllables, they might read entertaining prose or perform a simple mechanical task). The members of both groups are then tested for their retention of the original learning of task *A.* The time which has elapsed since the learning of *A* is exactly the same for the two groups; they differ only in the nature of the activity filling that time interval. To the extent that the retention of the experimental group is poorer than that of the control group, retroactive inhibition has taken place. The experimental paradigm of retroactive inhibition may be summarized as shown by the accompanying design.

Experimental Group:

Learn A	Learn B	Retention Test for A

Control Group:

Learn A	Rest	Retention Test for A
		Difference = Retroactive Inhibition

Retroactive inhibition, then, refers to the difference in degree of forgetting caused by the interpolation of a formal learning task as compared with a time interval free of formal learning activity. Even the control group is not entirely inactive but the amount of formal learning is minimized.

Retroactive inhibition is an example of negative transfer. The learning and retention of any one particular task or skill is never entirely unaffected by previous learning activities of the subject. Whenever the learning of a task, *B,* facilitates the learning or retention of another task, *A,* we speak of positive transfer. Whenever the

learning of a task, *B*, interferes with the learning or retention of another task, *A*, there is negative transfer (see Chapter 18). Retroactive inhibition is a special example of negative transfer because the learning of one task interferes with the *retention* of another.

The term, *retroactive inhibition,* should not be interpreted to imply backward action of one activity upon another. The interpolated activity (*B*) interferes with the results or "traces" of the previous activity (*A*). The term is, perhaps, unfortunate because of its possible misinterpretation in the sense of backward action; it is used merely to emphasize that it is the retention for an earlier learned task with which the interpolated activity interferes.

Determinants of Retroactive Inhibition

The basic paradigm and variations on it have been employed with a wide variety of stimulus materials, under many experimental conditions, using several methods of measuring degree of retention. From this wealth of experimentation, retroactive inhibition has emerged as a highly reliable and predictable phenomenon. At the same time, it has become clear that retroactive inhibition is not an all-or-none effect but varies in degree over a wide range as a function of a number of experimental variables. We turn next to a consideration of the known determinants of retroactive inhibition.

Method of Measurement. Retroactive inhibition is inferred from differences between retention scores resulting from differences in interpolated activity, as in the paradigm above. Amount of retroactive inhibition depends, therefore, on the particular method of measuring retention. Differences in active-recall scores are the most widely used, and are sensitive indicators of degree of retroactive inhibition. On the first retest trial after interpolated activity, the differences between recall scores of the experimental group and the control group are likely to be considerable. This difference, which may be converted into a percentage of loss, provides a reliable measure of retroactive inhibition. If the test trials are continued, i.e., if the two groups are required to relearn the material to the original criterion of mastery, the differences between the two groups will become less and less on successive trials. The effects of retroactive inhibition are rapidly dissipated during relearning, at least as far as they are measured by the numbers of items retained

Because of their initial advantage, however, the members of the control group will relearn the original material to mastery in fewer trials or less time than the members of the experimental group. Thus, differences in percent *saving* provide another measure of retroactive inhibition. Due to the transient nature of retroactive effects, this measure is likely to be less sensitive than active-recall scores obtained on the first test trial following interpolation.

Other retention measures—e.g., scores on tests of recognition and reconstruction—may also be used to gauge degree of retroactive inhibition although they are not as likely as the first two methods to yield fine discriminations between various conditions of learning and interpolation. In the case of both recognition and reconstruction tests, the presence of the correct items on the test may serve as a rehearsal which counteracts, at least in part, the effects of the interpolated activity.

Retroactive inhibition is most commonly measured by comparing the *amounts* retained under different conditions of interpolation. An interpolated activity may, however, not only decrease the sheer amount retained but also affect the efficiency of performance for those parts of the original learning which are correctly retained. Retroactive inhibition may, for example, result in reduced *speed* of performance. Fig. 93 demonstrates the effects of an interpolated activity on the speed (reaction time) with which paired associates are anticipated. Even when only correct responses are considered, retroactive inhibition manifests itself in consistent slowing of reaction times. Similarly, an interpolated activity reduces the *rate* at which correct responses are given on an active-recall test. Temporal indices such as these illustrate the far-reaching effects of an interpolated activity upon retention and the need for sensitive indices to gauge their full extent.

Similarity Between Original and Interpolated Activities. However measured, amount of retroactive inhibition is determined to a large extent by the degree of similarity between original and interpolated learning (*A* and *B* in our paradigm). If the original activity (*A*) is learning a list of nonsense syllables, an interpolated second list of nonsense syllables (*B*) will cause considerably more retroactive inhibition than the interpolated learning of a meaningful prose passage or of a motor skill. Two activities do not interfere

with each other too seriously if they involve entirely different types of responses. As the two activities become more similar and involve more responses in common, retroactive interference increases. Finally, the similarity between original learning and interpolated learning may become so great that the interpolated activity plays

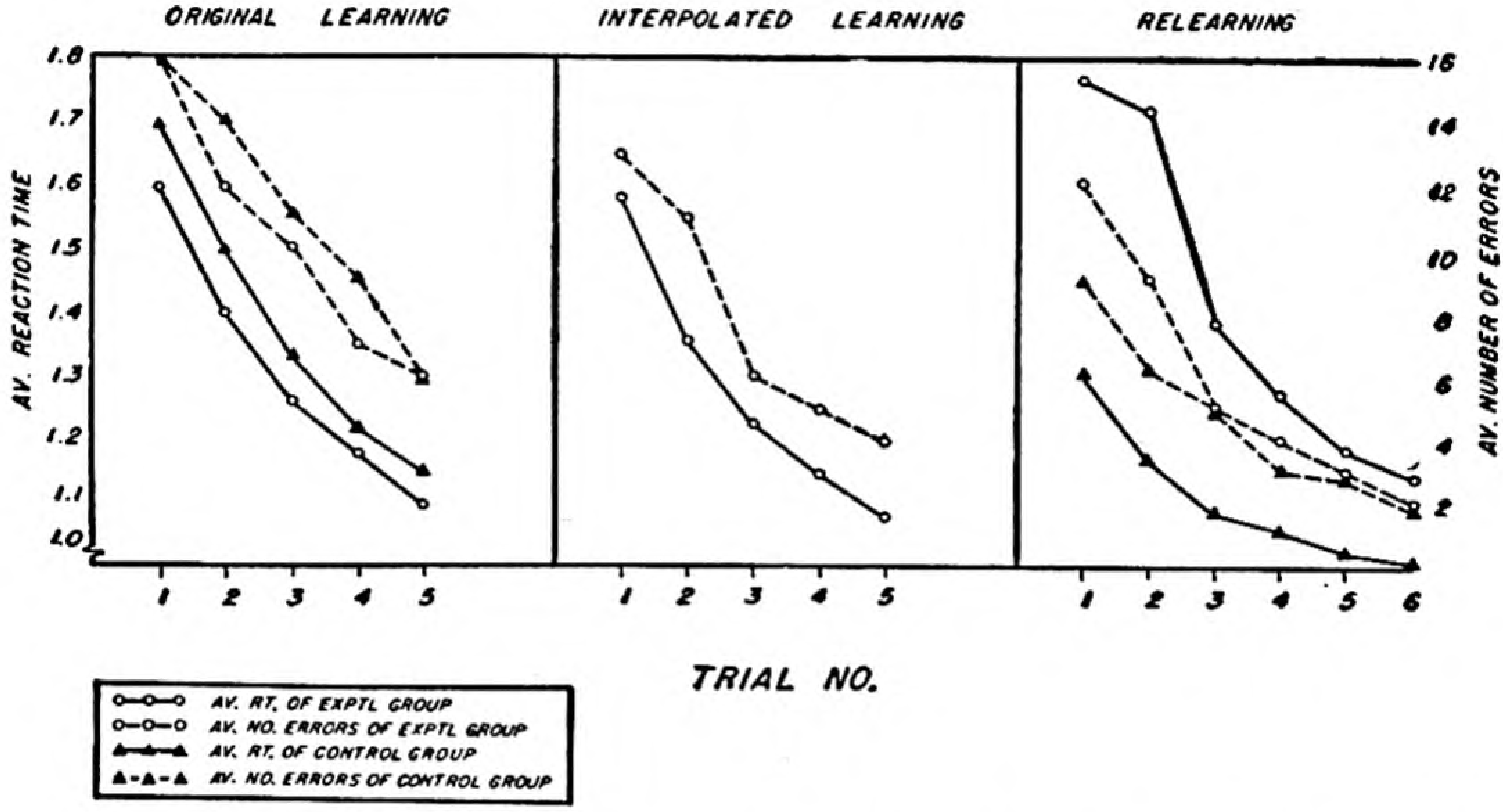


FIG. 93. The Measurement of Retroactive Inhibition. Both retention scores and reaction times may be used to gauge the effect of an interpolated activity. During the original learning, the errors decline and reaction times become faster. After interpolation, errors increase and reaction times become slower for the experimental group. The control group, which had not interpolated learning, shows no comparable increase in errors and reaction times. These data were obtained in an experiment using the method of paired associates. (From L. Postman and H. L. Kaplan, Reaction time as a measure of retroactive inhibition, *J. Exper. Psychol.*, 1947, 37:141, by permission of the journal and the American Psychological Association.)

the part of a rehearsal or practice trial for the original activity and serves to strengthen its retention rather than to inhibit it.

These considerations of the role of similarity of original and interpolated learning have led to the formulation of the *Skaggs-Robinson hypothesis* which can best be understood by reference to Fig. 94. The abscissa of Fig. 94 represents a hypothetical scale of similarity between original learning and interpolated learning, rang-

ing from complete identity (A) to complete unrelatedness (C). On the ordinate is plotted relative degree of retention following interpolation. According to the Skaggs-Robinson hypothesis, relative degree of retention is maximal when the two activities are identical. The transfer from one task to itself is positive. As similarity decreases, so does degree of recall until it reaches a minimum

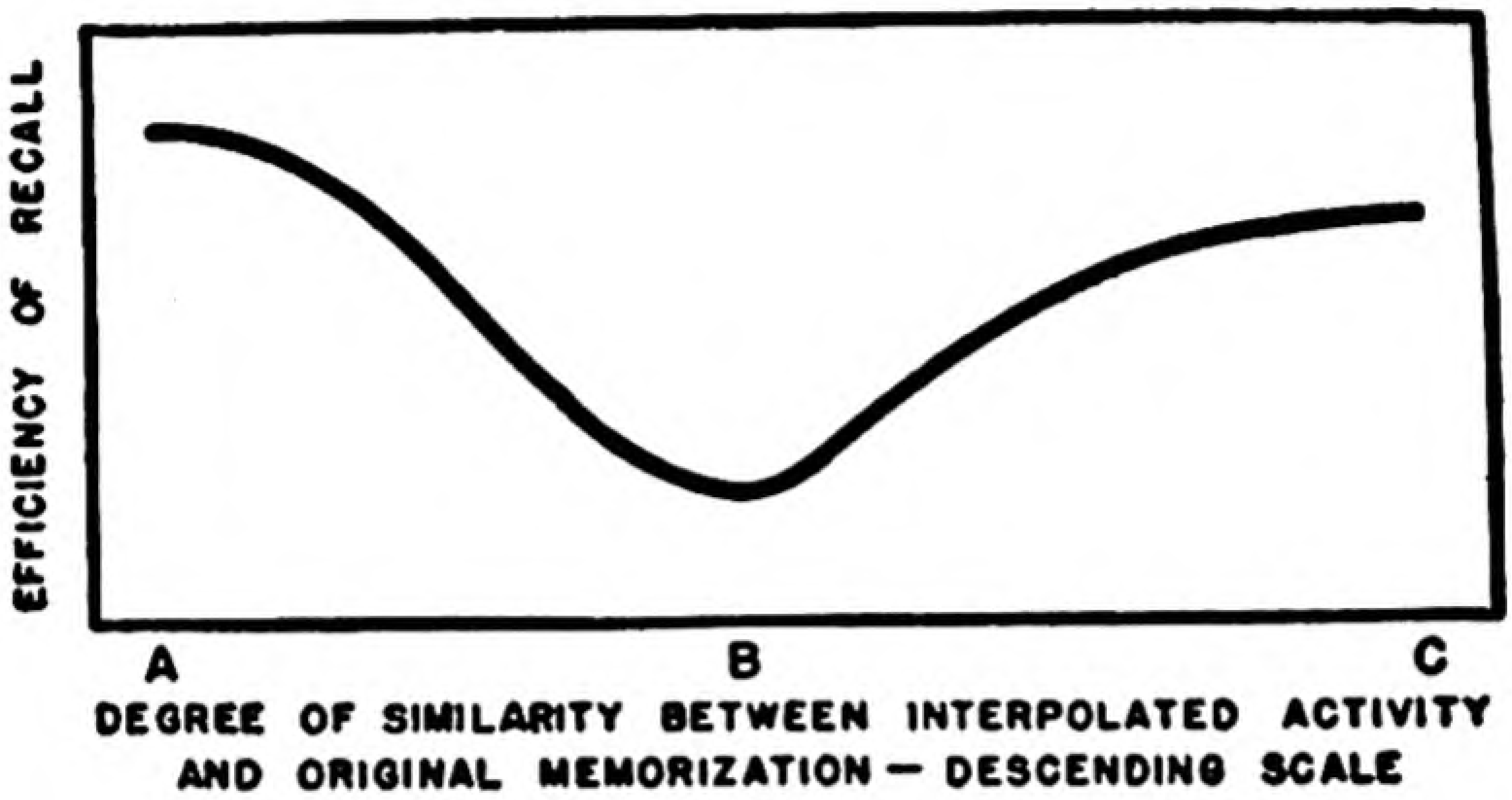


FIG. 94. This Hypothetical Curve Shows How Efficiency of Recall in a Retroactive Inhibition Experiment Varies as a Function of the Degree of Similarity Between Original and Interpolated Learning. When similarity is very great (A) or very small (C), efficiency is high, i.e., there is little interference. Maximum interference occurs with an intermediate degree of similarity (B). (From J. A. McGeoch, *The psychology of human learning*, 1942, p. 462, by permission of Longmans, Green & Co., Inc. After E. S. Robinson, the "similarity" factor in retroaction, *Amer. J. Psychol.*, 1927, 39:299, by permission of the journal.)

at an intermediate degree of similarity (B). At this point, the two activities involve a sufficient number of common responses to provide considerable opportunity for interference, but they are too far removed from identity to allow mutual facilitation. Two lists of different nonsense syllables are representative of this condition. At point B, then, the transfer from one task to the other is predominantly negative. As the similarity between the two tasks decreases even further, in the direction of complete unrelatedness, degree of retention rises until it reaches a second maximum at C. At C the

two tasks do not involve any common responses and there is little or no opportunity for interference. However, retention cannot be as high at *C* as it is at *A*, since complete identity of the original and interpolated tasks (*A*) means additional practice, whereas complete unrelatedness (*C*) simply implies a minimum of interference but no additional practice.

The Skaggs-Robinson hypothesis has received at least partial verification in experimental investigations of retroactive inhibition. Separate sections of the theoretical curve have been empirically approximated. Identity of original and interpolated learning does, of course, result in maximum retention. Any degree of similarity less than identity results in lower retention, increased retroactive inhibition (*A-B* section of Skaggs-Robinson curve). It has also been shown that over a considerable range, the amount of retroaction decreases, with decreasing degree of similarity (*B-C* section of Skaggs-Robinson curve). It has not been possible, at least thus far, to obtain empirically a curve corresponding to the theoretical Skaggs-Robinson function over its entire range. The curve shown in Fig. 94 has remained a theoretical model. The most serious obstacle to a conclusive test of the theory has been the difficulty of varying similarity of the interpolated activity continuously over a range as wide as that demanded by the theory.

We are thoroughly accustomed to think in terms of similarities and differences, but these concepts are difficult to measure and manipulate experimentally. Strictly speaking, similarity is a matter of the *learner's* perception of the stimulus materials: the experimenter must always guard against ascribing his own perception of similarity to the learner without independent test. Learning materials, moreover, may be similar or different in many respects: similarity can vary along many dimensions. Two learning tasks may be similar because they contain common (identical) elements or items, or because they are more or less alike in meaning. There may be similarity with respect to the set under which two tasks are performed or with respect to the operations and responses required of the learner. All such similarities have demonstrable effects on the degree of retroactive inhibition.

Similarity must be carefully defined in terms of the operations

which the experimenter uses to make the stimulus materials similar. A few examples will illustrate this point.

Similarity due to common elements. The similarity between two learning tasks may be due to the fact that they contain common elements or items. For example, it is possible to construct lists of nonsense syllables which overlap in structure. List *A* may contain the following syllables: HUJ, CEX, WAP; List *B*: HUZ, DEX, WIP. These syllables have two out of three letters in common. If similar syllables appear in corresponding positions in the two lists (as is the case in the example above), the degree of similarity between the two lists is enhanced. Two motor tasks may call for partially, but not entirely, identical movements; two mazes may have partially, but not entirely, identical paths. In all such cases, degree of similarity is defined by number of common stimulus elements.

Similarity of meaning. Stimulus materials may not be alike formally or physically and yet be perceived and responded to as similar because of common meaning and interpretation. Synonyms exemplify this dimension of similarity. When the original learning consists of a list of meaningful words, interpolation of a list of synonyms causes greater retroactive inhibition than the interpolation of a list of unrelated meaningful words. Such similarity can sometimes be defined *a priori* on the basis of dictionary meaning. An alternative and preferable procedure consists of obtaining ratings or judgments of similarity from a group of independent judges. In this manner, a rough scale of similarity can be constructed and amount of retroactive inhibition compared for different degrees of similarity. The rating procedure can be used for approximate scaling of similarity with a wide variety of stimulus materials, such as words and geometric designs.

Similarity of operations. Quite apart from the nature of the stimulus materials, tasks may be similar or different with respect to the types of responses or operations required of the subject. One list may, for example, be learned by the method of anticipation, the other, by the method of complete presentation. Similar or different motor responses, such as approach and withdrawal, may be called for in original and interpolated learning. In general, similarity of operations leads to increased retroactive interference.

Similarity of set. Similarity of operations is an example of a more

general dimension, similarity of *set*. The operation of a set results in *selective* response to the stimulus material and, hence, learning and retention. Amount of interference varies with the extent to which original and interpolated learning are carried on under the same or different selective sets. Similar selective sets maximize the responses common to the two tasks and increase retroactive inhibition. If one passage is memorized verbatim and another learned "for understanding," there is less interference than if both are learned under the same set. In general, the more different the sets under which the original and interpolated activity are learned, the less will be the retroactive interference. Whatever serves to reduce the responses common to the original and interpolated activity, decreases the amount of retroactive inhibition. The importance of "functional isolation" of the two activities is dramatically illustrated with the aid of hypnosis. If one of the tasks (either original or interpolated) is learned under hypnosis while the other is learned in a state of wakefulness, the amount of retroactive inhibition is reduced as compared with normal conditions (both tasks learned in a waking state).

To summarize our discussion of the factor of similarity: whatever leads to establishment of incompatible or competing responses for the original and interpolated activities, increases retroactive inhibition. In practice, similarity between the two activities (short of identity) with respect to materials, methods and sets will, over a wide range, result in decreased retention of the original task.

Amount and Strength of Original and Interpolated Learning. Holding the nature of the tasks and the subject's set constant, amount of retroactive inhibition depends on the strength to which original and interpolated learning are carried, as well as on the amount of the two activities.

The better learned (or the more overlearned) the original activity is, the less susceptible it is to retroactive inhibition. A strongly established response cannot be easily disturbed. A simple example will illustrate this point. We continually (and often to our embarrassment) forget the names of new acquaintances, probably because there is considerable retroactive inhibition among poorly learned names or name-face associations. No matter how many new names we have to learn, however, there is no retroactive inhibition

for our own names or those of our closest friends—these responses are too well established. In experimental practice, the greater the frequency of repetitions of the original material or the more exacting the criterion to which it is carried, the less is the detrimental effect of an interpolated activity.

The interpolated activity must be learned to at least a minimum strength in order to be an effective inhibiter. If the interpolated task is greatly overlearned, however, it cannot be easily "confused" with the original activity, and amount of retroactive inhibition diminishes. If amount of retroactive inhibition is plotted against degree of interpolated learning (other things being equal), we obtain a function which rises to a maximum and then declines as the strength of the interpolated learning exceeds a certain value.

In addition to the strength of the original and interpolated learning taken separately, it is necessary to consider their strength relative to each other. If one is extremely strong as compared with the other, there is little mutual interference. Maximum retroaction occurs when original and interpolated learning are approximately equal in strength. If both tasks have been learned to a moderate degree, considerable interference effects may be expected.

Temporal Point of Interpolation. The interpolated activity may be introduced at varying time intervals after the end of the original learning. An interpolated task may effectively reduce the efficiency of recall even if it is introduced a long time after the end of practice. In all probability, an interpolation will have some detrimental effect so long as there is retention for the original activity.

Amount of retroaction may vary with the exact point in time at which the interpolated activity is introduced. Let us assume a constant time interval between the end of practice and the test of retention. An interpolated activity may be introduced at various intermediate points. The experimental work on the effect of this variable has been neither extensive nor conclusive. Nevertheless, the evidence points to two temporal loci of interpolation which produce maximum interference: (1) interpolation shortly after the end of practice, and (2) interpolation shortly before the test of retention. Other times of interpolation yield intermediate degrees of retroaction. It is possible that the process of retroactive inhibition may

be due to two factors. One of these factors may have maximum effectiveness immediately after the original learning, the other, shortly before recall. In this way the existence of the two maximally effective points of interpolation could be explained. It is to a consideration of such a two-factor theory that we turn next.

A Two-Factor Theory of Retroactive Inhibition. Consider a typical experiment on retroactive inhibition: original learning—test—interpolated activity—retest. If we analyze carefully the records obtained on the retest, we frequently find *intrusions* from the interpolated activity. For example, syllables or words from the interpolated list will appear among the items recalled from the original list. The occurrence of such overt intrusions suggests that retroactive inhibition is due, at least in part, to competition among responses established during original and interpolated learning. Such competition can best be demonstrated when the following experimental procedure is used:

Original Learning	Interpolated Learning	Retest
A-B	A-K	A-B

The experiment is of the paired-associate type. The left-hand member of the pair (A) is the same during original and interpolated learning. The right-hand member varies (B and K). In this manner, two different incompatible responses, B and K, are connected with A. During the retest, these two responses compete with each other, and there may be frequent overt intrusions of K in the place of B.

Overt intrusions, however, do not account for the total amount of loss due to the interpolated activity. For this reason, another factor—sometimes called "unlearning"—has been postulated. During the interpolated learning, some of the original responses may be so much weakened or inhibited that they are made completely unavailable for recall, i.e., they are "unlearned." Such unlearning may have some of the properties of the extinction of conditioned responses. During the interpolated activity, the original connections, e.g., *A-B* in our paradigm, fail to receive any reinforcement, since the correct connection now is *A-K*, and, failing of reinforcement, become weaker and weaker until they are unlearned.

According to the two-factor theory, then, the process of retroactive inhibition may be conceptualized in terms of two factors—

unlearning and competition among responses. Unlearning consists of the weakening of the original responses so that they are no longer available for recall. Competition of responses results from the fact that two incompatible response tendencies are acquired during original and interpolated learning. These response tendencies compete with each other during the retention test.

Retention as a Function of the Test Situation

It would be an error to localize the determinants of forgetting exclusively in the conditions of original learning and in the time interval between the end of practice and recall. The conditions at the time of the retention test are of critical importance. We have already emphasized that degree of retention is a function of the particular test used. With any given test, maximum retention is obtained if the original training procedure has involved preparation for the type of test by which retention is measured.

The context in which retention is measured is important. The learner forms associations not only among the items which he explicitly attempts to master but also between the performance of his task and a definite environmental context. Physical environment, time of day, the personality of the experimenter, the presence or absence of an audience—all these form a background of stimulation with which the performance of a specific task is more or less strongly associated during the practice period. When retention is tested, an environmental context identical, or at least very similar, to the practice situation provides strong "behavioral support" to the learner. There is good experimental evidence that alterations in environmental context impair the efficiency of retention. Those who have had occasion to perform a familiar task in an unfamiliar environment can readily verify this fact.

In some ways, the effects of altered stimulus context are related to retroactive inhibition. A new or changed environment requires new adjustments, new associations, as it were, between situation and the performance of the task. These new associations interfere with those formed during the original practice period. Such interference is not very serious if the task has been overlearned to a degree which makes it independent of supports from the environmental stimulus context. In this respect, too, there is a parallel to retroactive inhibi-

tion, for highly overlearned activities cannot be affected appreciably by interpolated tasks. Retroactive inhibition and the detrimental effects of altered stimulus contexts are seen to be continuous and governed by similar principles. In both cases, the retention of the original activity is disrupted by the intrusion of new associations.

Reminiscence

The Phenomenon of Reminiscence. In general, degree of retention continually diminishes as the time interval between the end of practice and the test of retention increases. Under certain con-

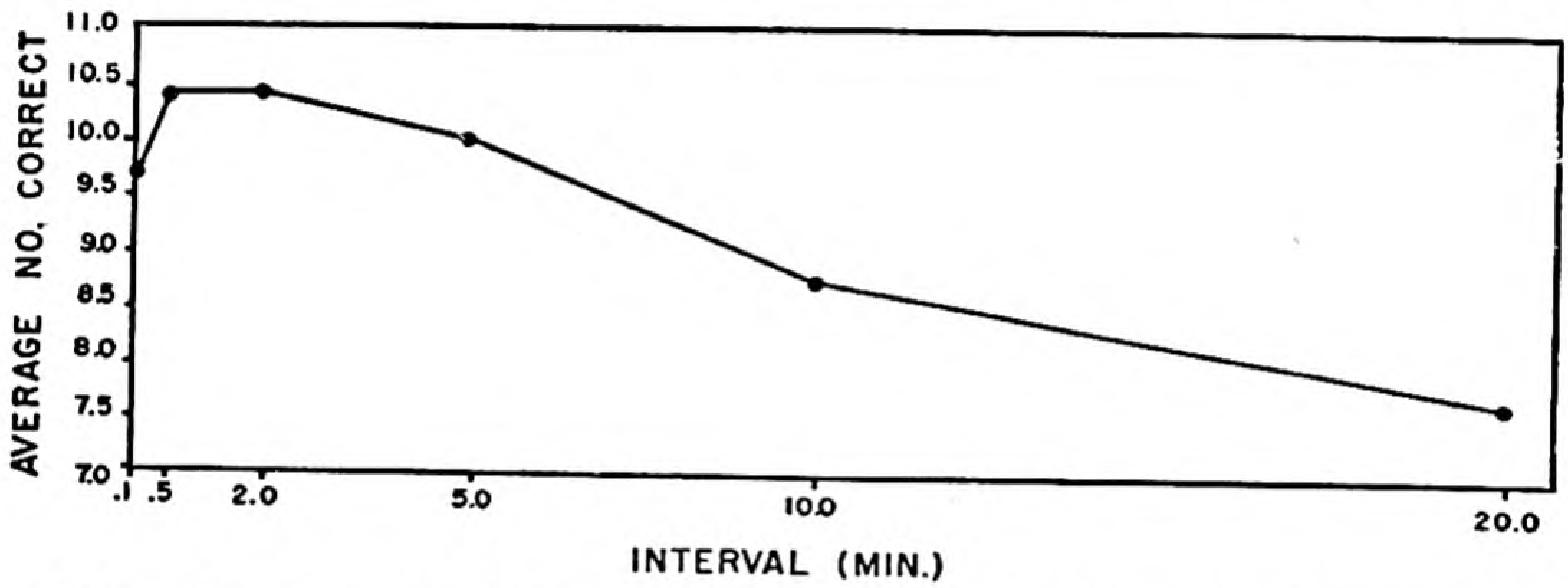


Fig. 95. Retention Curve Showing the Reminiscence Phenomenon. (After L. B. Ward, Reminiscence and rote learning, *Psychol. Monogr.*, 1937, *49*, No. 220, p. 30, by permission of the journal and the American Psychological Association.)

ditions, however, the curve of retention shows a reversal: amount retained at first increases as a function of time and then decreases in the usual manner. Such a temporary improvement in performance occurring without the benefit of practice is known as *reminiscence.* Fig. 95 presents a retention curve showing the reminiscence phenomenon. The fact of reminiscence underscores the point that time in and of itself is not a condition of forgetting, for processes in time may lead to an increase rather than a decrease in retention.

Experimental Demonstration of Reminiscence

Two procedures have been used to demonstrate reminiscence.

Successsive Test Performances. Subjects are given a certain amount of practice and their retention is tested. Some time later

(the intervals are usually on the order of a few days), their retention is again tested, without any intervening practice. If performance on the second test is better than on the first, reminiscence is said to have occurred. Amount of reminiscence is defined by the percent improvement on the second test as compared with the first.

This procedure is open to serious methodological criticism. First and most important, it is difficult, if not impossible, to control rehearsal during the period between the two tests. There is no formal practice, but informal rehearsal by the subject cannot be prevented, especially when the interval between the two tests is a day or more. The only evidence we can obtain regarding review between tests is the subjects' own reports. It is true that amount of rehearsal *as reported by the subjects* seems to be uncorrelated with degree of reminiscence. Those who report that they did not rehearse the learning material between tests often show considerable reminiscence. Subjects' testimonies, however, are at best unreliable bases for evaluating the effects of informal review between tests.

There is a second difficulty. The first of the two tests is not only a test but also in many ways a practice trial. Performance of the learning task strengthens the associations established during the practice period. The second test may thus benefit from the opportunity for practice provided by the first test. Furthermore, new items may emerge on the second test, partly because the first test has strengthened the organization of the material, and partly because items near the threshold will sometimes be available for recall and sometimes not. Since reminiscence is defined as *improvement in performance without benefit of practice,* effects due to practice on the first test cannot be considered as true reminiscence.

The two basic difficulties in the type of reminiscence experiment under consideration thus are (1) informal review between successive tests, and (2) opportunity for practice provided by the first test. These difficulties are adequately met by another type of reminiscence experiment which has been widely used in recent years.

Use of Equated Groups. Instead of testing the *same* subjects on two successive occasions, we use two equated groups. Both groups learn the material to a given degree of mastery, as measured on a criterion test. One group receives a retest immediately

following the criterion test. The other group receives a retest some time after the criterion test. Typically, short intervals of the order of a few minutes are used. During this interval, an attempt is usually made to control the subjects' activity in order to prevent or minimize rehearsal. For example, if the learning material consists of verbal items, the subjects may be engaged in naming colors or other tasks which are as unrelated as possible to the activity practiced. Comparison of the performances of the two groups on the retest yields a measure of reminiscence. If the retest following the criterion test after an interval of time yields better results than the immediate retest, reminiscence has occurred. The percent difference between the two retests yields a measure of reminiscence.

This procedure minimizes the problem of informal rehearsal by using short time intervals during which the subjects' activity can be reasonably well controlled. The problem of practice effects is held constant since both groups receive a test and a retest. The better equated the two groups are, the more valid the conclusions regarding reminiscence which can be based on such an experimental procedure, since differences between them can then be ascribed to the difference in time interval between test and retest.

Determinants of Reminiscence

At first glance, the fact of reminiscence appears most improbable. Why should performance improve rather than decline as a function of time without intervening practice? Again, time in and of itself does not do anything. The time interval between test and retest does, however, allow inhibitory effects which accumulate during practice to be dissipated. In serial learning, for example, remote associations are formed among nonadjacent items, leading to anticipatory and perseverative errors (for a full discussion of intraserial inhibition, see pp. 323-326). In motor learning, inadequate and wrong responses are acquired along with the correct ones. Wrong and inadequate responses are reinforced less frequently and are, therefore, weaker than correct responses. During the time interval following the end of practice, such weak responses are forgotten more quickly than the stronger correct responses. Such *differential forgetting* may account, at least in part, for the improve-

ment of performance following a time interval. Thus, reminiscence and the beneficial effects of distributed practice (see pp. 332-336) are in all likelihood closely related. In both cases, dissipation of inhibitory connections or differential forgetting leads to improved performance.

If reminiscence is due primarily to the dissipation of inhibition, reminiscence should be most pronounced if a considerable amount of inhibition is built up during the practice period. There is good experimental evidence that such is indeed the case.

Spacing of Practice. Massed periods of concentrated practice with only short rest intervals lead to an accumulation of inhibitory effects and do not allow them to be dissipated. With distributed practice, on the other hand, differential forgetting can take place. Reminiscence effects are more pronounced after massed than after distributed practice, supporting the view that dissipation of inhibition is at least in part responsible for the phenomenon.

Effects of Serial Position. In serial learning, maximum inhibition effects occur in the middle of the list, which is spanned by the greatest number of remote associations (cf. p. 325). As the theory would lead us to believe, the middle items of a list show greater reminiscence than items at the beginning and end of a list.

Rate of Presentation of Items. Closely related to the effects of spacing is the dependence of reminiscence on the rate at which items are exposed during the practice period. A fast rate of exposure—item following upon item in quick succession—leads to more reminiscence than does a slow rate of exposure. Presumably, a fast rate of exposure leads to a greater accumulation of inhibitory effects, which are dissipated during the time interval following the end of practice.

Method of Practice. When the conditions of learning do not favor the accumulation of inhibitory effects, reminiscence may fail to occur. For example, when verbal items are learned by the method of paired associates, the pairs are presented in random order and remote associations are minimized. Again, in accordance with a differential forgetting theory, learning by paired associates fails to show reminiscence, whereas learning by the anticipation method does.

In spite of such supporting evidence, the inhibitory or differential forgetting theory of reminiscence cannot as yet be regarded as established fact. There are results on record showing failure of reminiscence even when intraserial inhibition (as shown, for example, by serial position effects) is present. Whether factors other than the dissipation of inhibition may be responsible for reminiscence must remain an open question.

Degree of Learning. Reminiscence, like all measures of retention, varies with the degree of mastery to which learning has been carried. If learning is poor, there is little reminiscence, for few associations are strong enough to be elicited on the retest. On the other hand, if the degree of learning is great, there is little room for improvement or reminiscence. It is thus for intermediate degrees of learning that the number of items reminisced is greatest. As far as relative reminiscence is concerned, i.e., the percentage of the number of items in the total list which is reminisced, it tends to vary inversely with degree of learning.

Measurement of Reminiscence. Reminiscence is defined as improvement in retention without benefit of practice. Comparison of retention scores obtained immediately and at different time intervals after the end of practice thus provides a measure of reminiscence. Before designating improvement in retention as reminiscence, the statistical significance of the increment in retention must be tested.

In actual experimental practice, the comparison of average retention scores has been found unsatisfactory in the measurement of reminiscence. The difficulty is that in a group of subjects, some will show reminiscence while others will not. Reminiscence by some and forgetting by others will cancel each other when the results are averaged. For this reason, experimental reports often state the percentage of subjects who show reminiscence. If a larger percentage of experimental subjects shows reminiscence than can be expected on the basis of chance variation, there is evidence of reminiscence.

Reminiscence can also be gauged by an analysis of individual items. We can tabulate the number of items which persist from first test to retest. It is the new items which failed to appear on the first test but appear on the retest which constitute reminiscence. The

results of such an item analysis appear in Fig. 96. The number of items recalled on the test immediately following practice constitutes the 100-percent level and subsequent recalls are expressed as per-

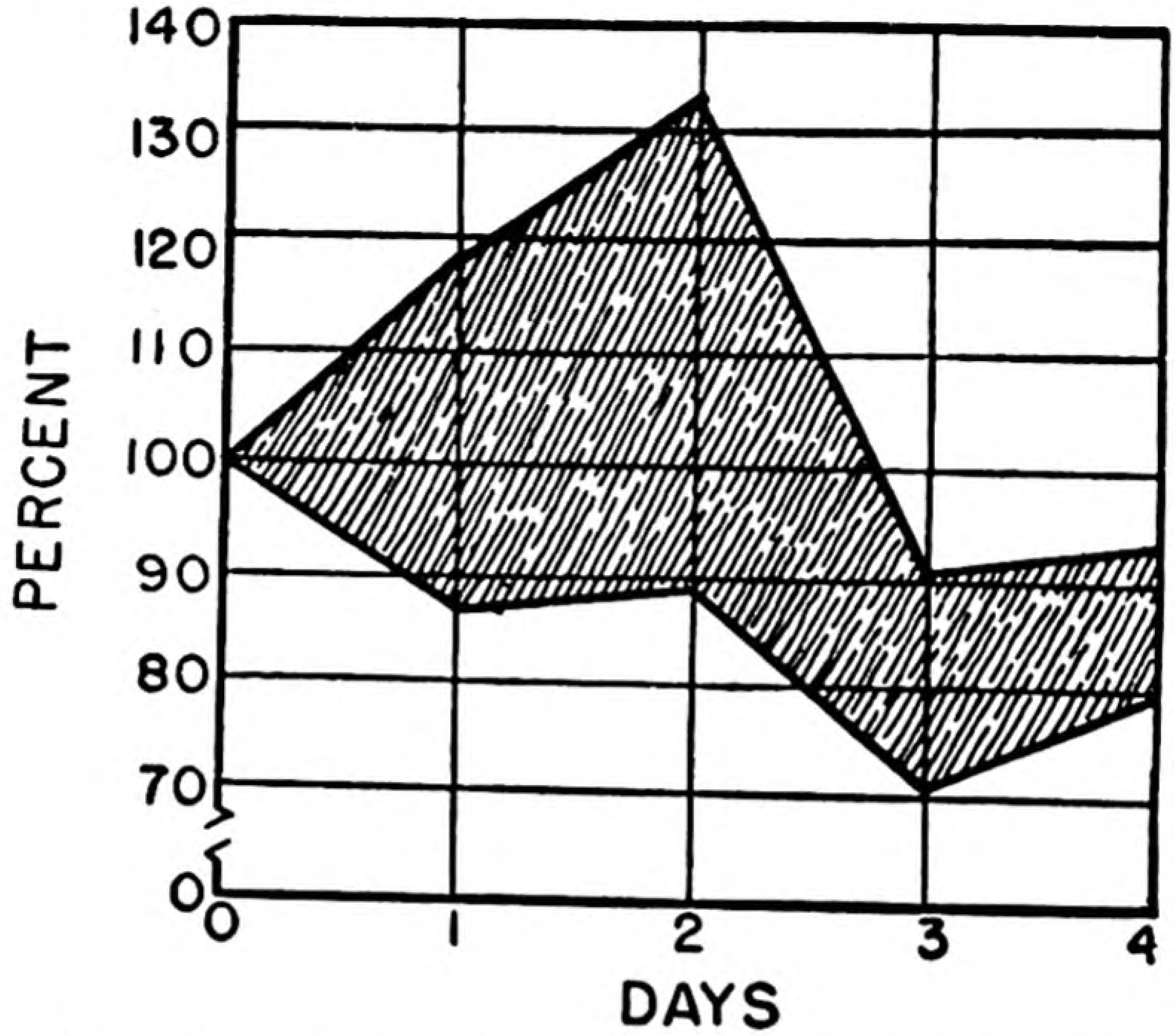


FIG. 96. The Measurement of Reminiscence. This curve shows to what extent reminiscence is due to the appearance of new items on successive tests. The upper curve shows the total retention scores at varying time intervals after the end of practice. The lower curve shows the retention for those items which were recalled on the test immediately following learning. While some of the originally recalled items are forgotten, new items come in, yielding recall scores about 100 percent. The shaded area represents the reminiscence effect due to the appearance of new items without practice. (After P. B. Ballard, Oblivescence and reminiscence, *Brit. J. Psychol. Monogr. Suppl.*, 1913, *1*:22, by permission of the journal.)

centages to that base. The lower curve in Fig. 96 shows the percentage of items persisting from test to retest; the upper curve shows the *total* retention scores obtained at varying time intervals after the end of practice. While some of the items originally recalled

are forgotten, new items come in and cause the total recall curve to rise above 100. The shaded area, therefore, represents the reminiscence effect—the new items which are recalled without intervening practice.

Experiment XXII
Retroactive Inhibition

Purpose. To demonstrate the phenomenon of retroactive inhibition and to study one of its most important determinants—similarity of original and interpolated learning.

Materials. A list of 12 two-syllable adjectives (List A), a second list of 12 two-syllable adjectives (List B), which are synonyms of the adjectives in List A, and a list of 12 three-place numbers (List C) are required. A memory drum is used for the exposure of the stimulus items. Prepared score sheets as shown ~~on p. 345~~ are required for recording the results.

Procedure. The subjects are divided into three groups: two Experimental Groups—the *Similar* Group and the *Different* Group—and a Control Group.

The procedures used with the three groups are as follows:

1. *Similar Group.* The members of this group learn List A by the method of anticipation to the criterion, one perfect repetition. Immediately after reaching this criterion, these subjects receive six trials with List B.[3] They are then given a rest period of 5 minutes during which they read some light prose or engage in some other activity unrelated to the learning task. It is important to control the activity of the subjects during the rest period in order to prevent review. At the end of the rest period, List A is relearned to a criterion of one perfect repetition.
2. *Different Group.* The procedure for this group is identical with that for the *Similar* Group except for the interpolated activity. In this case, List C is interpolated for six trials.
3. *Control Group.* The members of this group learn List A to criterion, rest (i.e., engage in an activity unrelated to the learning task) and then relearn List A to criterion. It is naturally important that the length of the rest period be equal to the amount of time which elapses between the end of learning *A* and the relearning of *A* in the case of the experimental groups.

[3] A constant number of practice trials is used for the interpolated condition in order to make possible equation of the retention intervals of the three groups.

The experimental design illustrates the paradigm conventionally used in the study of retroactive inhibition. It may be summarized as follows:

Control Group:
Learn *A* — Rest — Relearn *A*
Similar Group:
Learn *A* — Learn *B* — Rest — Relearn *A*
Different Group:
Learn *A* — Learn *C* — Rest — Relearn *A*

The Control Group, then, rests while the experimental groups practice an interpolated activity. The two experimental groups differ from each other in the nature of the interpolated activity. For the *Similar* Group, the original and the interpolated activities are very much alike, viz., one list consists of synonyms of the other. For the *Different* Group, the two activities are much less similar, a list of adjectives and a list of three-place numbers.

Treatment of Results. Retroactive inhibition is defined as decrement in retention due to an interpolated activity. To demonstrate this effect, we compare the relearning performance of the Control Group with that of the Experimental Groups.

Two measures of retention may be used: (1) the number of items correctly anticipated on the first relearning trial, and (2) the number of trials required for relearning to criterion. These two measures are computed for the Control Group and the Experimental Groups (at this stage of the analysis, the data for the two experimental groups may be pooled). If the performance of the Experimental Groups is significantly poorer than that of the Control Group, retroactive inhibition has been demonstrated.

To evaluate the effect of similarity of original and interpolated activity, relearning performances of the two experimental groups are compared with each other. The same two measures of retention—number of items correctly anticipated on the first relearning trial and number of trials required to relearn to criterion—are used. Comparison of the two experimental groups on these measures will show whether or not, and to what extent, the degree of similarity of original and interpolated learning has influenced the amount of retroactive inhibition.

The results of this experiment also provide data for plotting serial position curves.

Experiment XXIII
Retention for Completed and Interrupted Tasks[4]

Purpose. To test the hypothesis that, other things being equal, interrupted tasks are retained better than completed tasks.

Materials. For the purposes of this experiment, it is necessary for the subject to perform a large number of tasks, say, twenty. The tasks should meet two requirements: (1) it should be possible to complete each of them within a fairly short period of time, such as 3 to 5 minutes; (2) it should also be possible to interrupt these tasks at any time prior to completion. They should not, therefore, be problems which the subject can solve by a verbal formula or by recognition of a principle. For these reasons, manual tasks, such as card sorting, tweezer-dexterity tests, cancellation of letters, arrangements of blocks, coding exercises, etc., are suitable. For a list of twenty tasks successfully used in this type of experiment, see the monograph of Martin.[5]

Experimental Design. The problem of this experiment is to determine whether interrupted tasks are remembered better than completed ones. The subjects are, therefore, allowed to finish half the tasks and are interrupted in the performance of the other half. We must make sure, however, that whatever differences in retention are found may be reasonably ascribed to the factor of interruption rather than to differences in the ease with which various tasks can be remembered. (One task may be more familiar, more vivid, more interesting than another and be remembered better for these reasons.)

For purposes of controlling possible differences in the ease of remembering the various tasks, we divide the subjects into two groups. The first group is allowed to complete all the even-numbered tasks (a number having been assigned to each of the tasks) and is interrupted in the performance of the odd-numbered ones; the second group is allowed to complete the odd-numbered tasks and interrupted in the performance of the even-numbered ones. Of course, the order in which the tasks are presented should be randomized, and uniform alternation of completed and interrupted tasks should be avoided.

Procedure. It is essential that the subject be in ignorance of the purpose of the experiment. The series of tasks is presented to the subject under some plausible pretext, e.g., that his coöperation is needed in the

[4] Only "naïve" subjects can be used in this experiment.

[5] J. R. Martin, Reminiscence and Gestalt theory, *Psychol. Monogr.*, 1940, **52**, No. 235.

standardization of a series of tests. Above all, the subject should not be led to suspect that his retention for the tasks will be tested.

The subject performs the entire series of tasks during one session. The tasks are presented to him, one by one, with appropriate instructions. If a task has been earmarked for interruption, the experimenter watches the subject's performance closely and interrupts his work before completion. The experimenter informs the subject that he has reached the time limit for this test and, after a brief rest period, introduces him to the next task. If, on the other hand, a task is scheduled for completion, the subject is, of course, allowed all the time he needs to carry it out. It is desirable that the average time spent on completed and interrupted tasks be equal.

A few (say, 3) minutes after the last task has been completed, the experimenter hands the subject a piece of paper and requests him to write down all the tasks that he remembers doing during the preceding session. The subject is allowed to continue until he fails to remember anything for a predetermined period (say, 5 minutes).

Treatment of Results. The experimenter tabulates the number of completed (C) and interrupted (I) tasks recalled by the subjects. He then divides the number of incomplete by the number of completed tasks, thus determining the I/C ratio (see p. 367). If the ratio is greater than one, interrupted tasks are remembered better than completed ones. Conversely, superior retention of completed tasks would yield a ratio smaller than one. The significance of the difference between the average numbers of completed and interrupted tasks remembered must be tested.

General References

Ebbinghaus, H., *Memory: a contribution to experimental psychology.* Trans. by H. A. Ruger and C. E. Bussenius, New York: Teachers College, Columbia University, 1913.

Katona, G., *Organizing and memorizing: studies in the psychology of learning and teaching,* New York: Columbia University Press, 1940.

Koffka, K., *Principles of Gestalt psychology,* New York: Harcourt, Brace and Co., 1935, Chap. 11.

McGeoch, J. A., *The psychology of human learning,* New York: Longmans, Green and Co., 1942.

Woodworth, R. S., *Experimental psychology,* New York: Henry Holt and Co., 1938, Chaps. 2, 3, 4, and 9.

Measurement of Retention

See General References, especially Ebbinghaus, McGeoch, and Woodworth

Anastasi, A. Further studies on the memory factor, *Arch. Psychol.*, 1932, No. 142.

Andrew, D. M., and Bird, C., A comparison of two new-type questions: recall and recognition, *J. Educ. Psychol.*, 1938, *29*:175-193.

Luh, C. W., The conditions of retention, *Psychol. Monogr.*, 1922, *31*, No. 142.

Postman, L., and Kaplan, H. L., Reaction time as a measure of retroactive inhibition, *J. Exper. Psychol.*, 1947, *37*:136-145.

Raffel, G., The effect of recall on forgetting, *J. Exper. Psychol.*, 1934, *17*:828-838.

Determinants of the Rate of Forgetting

Degree of Learning

Krueger, W. C. F., The effect of overlearning on retention, *J. Exper. Psychol.*, 1929, *12*:71-78.

Krueger, W. C. F., Further studies in overlearning, *J. Exper. Psychol.*, 1930, *13*:152-163.

Youtz, A. C., An experimental evaluation of Jost's laws, *Psychol. Monogr.*, 1941, *53*, No. 238.

Distribution of Practice

Cain, L. F., and Willey, R. de V., The effect of spaced learning on the curve of retention, *J. Exper. Psychol.*, 1939, *25*:209-214.

Hovland, C. I., Experimental studies in rote-learning theory: VI. Comparison of retention following learning to same criterion by massed and distributed practice, *J. Exper. Psychol.*, 1940, *26*:568-587.

Amount of Material

Robinson, E. S., and Heron, W. T., Results of variations in length of memorized material, *J. Exper. Psychol.*, 1922, *5*:428-448.

Robinson, E. S., and Darrow, C. W., Effect of length of list upon memory for numbers, *Amer. J. Psychol.*, 1924, *35*:235-243.

Nature of Learning Task

Buxton, C. E., and Newman, E. B., The forgetting of 'crowded' and 'isolated' materials, *J. Exper. Psychol.*, 1940, *26*:180-198.

McGeoch, J. A., The comparative retention values of a maze habit, of

nonsense syllables, and of rational learning, *J. Exper. Psychol.*, 1932, *15*:662-680.

Restorff, H. v., Über die Wirkung von Bereichsbildungen im Spurenfeld (Analyse von Vorgängen im Spurenfeld), *Psychol. Forsch.*, 1933, *18*:299-342. (English summary in K. Koffka, *Principles of Gestalt psychology*, New York: Harcourt, Brace and Co., 1935, pp. 481-489.)

Motivational Determinants

Interests, Attitudes and Values

Clark, K. B., Some factors influencing the remembering of prose material, *Arch. Psychol.*, 1940, No. 253.

Edwards, A. L., Political frames of reference as a factor influencing recognition, *J. Abn. Soc. Psychol.*, 1941, *36*:34-50.

Levine, J. M., and Murphy, G., The learning and forgetting of controversial material, *J. Abn. Soc. Psychol.*, 1943, *38*:507-517.

Seeleman, V., The influence of attitude upon the remembering of pictorial material, *Arch. Psychol.*, 1940, No. 258.

Sharp, A. A., An experimental test of Freud's doctrine of the relation of hedonic tone to memory revival, *J. Exper. Psychol.*, 1938, *22*:395-418.

Watson, W. S., and Hartmann, G. W., The rigidity of a basic attitudinal frame, *J. Abn. Soc. Psychol.*, 1939, *34*:314-335.

Completed and Interrupted Tasks

Martin, J. R., Reminiscence and Gestalt theory, *Psychol. Monogr.*, 1940, *52*, No. 235.

McKinney, F., Studies in the retention of interrupted learning activities, *J. Comp. Psychol.*, 1935, *19*:265-296.

Rosenzweig, S., Preferences in the repetition of successful and unsuccessful activities as a function of age and personality, *J. Genet. Psychol.*, 1933, *42*:423-440.

Zeigarnik, B., Das Behalten erledigter und unerledigter Handlungen, *Psychol. Forsch.*, 1927, *9*:1-85.

Retroactive Inhibition

General References

Britt, S. H., Retroactive inhibition: a review of the literature, *Psychol. Bull.*, 1935, *32*:381-440.

Britt, S. H., Theories of retroactive inhibition, *Psychol. Rev.*, 1936, *43*:207-216.

Swenson, E. J., *Retroactive inhibition: a review of the literature*, Minneapolis: University of Minnesota Press, 1941.

Retention After Sleep and Waking

Jenkins, J. G., and Dallenbach, K. M., Oblivescence during sleep and waking, *Amer. J. Psychol.*, 1924, *35*:605-612.

Minami, H., and Dallenbach, K. M., The effect of activity upon learning and retention in the cockroach, *Periplaneta americana*, *Amer. J. Psychol.*, 1946, *59*:1-58.

Van Ormer, E. B., Retention after intervals of sleep and of waking, *Arch. Psychol.*, 1932, No. 137.

Similarity Between Original and Interpolated Activities

Gibson, E. J., and Gibson, J. J., Retention and the interpolated task, *Amer. J. Psychol.*, 1934, *46*:603-610.

Harden, L. M., The effect of emotional reactions upon retention, *J. Gen. Psychol.*, 1930, *3*:197-221.

Kennelly, T. W., The role of similarity in retroactive inhibition, *Arch. Psychol.*, 1941, No. 260.

McGeoch, J. A., and McDonald, W. T., Meaningful relation and retroactive inhibition, *Amer. J. Psychol.*, 1931, *43*:579-588.

Robinson, E. S., Some factors determining the degree of retroactive inhibition, *Psychol. Monogr.*, 1920, *28*, No. 128.

Robinson, E. S., The 'similarity' factor in retroaction, *Amer. J. Psychol.*, 1927, *39*:297-312.

Skaggs, E. B., Further studies in retroactive inhibition, *Psychol. Monogr.*, 1925, *34*, No. 161.

Amount and Strength of Original and Interpolated Learning

McGeoch, J. A., Studies in retroactive inhibition: VII. Retroactive inhibition as a function of the length and frequency of presentation of the interpolated lists, *J. Exper. Psychol.*, 1936, *19*:674-693.

Melton, A. W., and Irwin, J. McQ., The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses, *Amer. J. Psychol.*, 1940, *53*:173-203.

Robinson, E. S., and Darrow, C. W., Effect of length of list upon memory for numbers, *Amer. J. Psychol.*, 1924, *35*:235-243.

Robinson, E. S., and Heron, W. T., Results of variations in length of memorized material, *J. Exper. Psychol.*, 1922, *5*:428-448.

Twining, P. E., The relative importance of intervening activity and lapse of time in the production of forgetting, *J. Exper. Psychol.*, 1940, *26*:483-501.

Underwood, B. J., The effect of successive interpolations on retroactive and proactive inhibition, *Psychol. Monogr.*, 1945, *59*, No. 3.

Temporal Point of Interpolation

Köhler, W., *Dynamics in psychology,* New York: Liveright Publishing Corp., 1940, Chap. 3.

Postman, L., and Alper, T. G., Retroactive inhibition as a function of the time of interpolation of the inhibitor between learning and recall, *Amer. J. Psychol.,* 1946, *59*:439-449.

Two-Factor Theory

Melton, A. W., and Von Lackum, W. J., Retroactive and proactive inhibition in retention: evidence for a two-factor theory of retroactive inhibition, *Amer. J. Psychol.,* 1941, *54*:157-173.

RETENTION AS A FUNCTION OF THE TEST SITUATION

Dulsky, S. G., The effect of a change of background on recall and relearning, *J. Exper. Psychol.,* 1935, *18*:725-740.

Pan, S., The influence of context upon learning and recall, *J. Exper. Psychol.,* 1926, *9*:468-491.

REMINISCENCE

Ballard, P. B., Oblivescence and reminiscence, *Brit. J. Psychol. Monogr. Suppl.,* 1913, *1,* No. 12.

Buxton, C. E., The status of research in reminiscence, *Psychol. Bull.,* 1943, *40*:313-340.

Hovland, C. I., Experimental studies in rote-learning theory: I. Reminiscence following learning by massed and by distributed practice, *J. Exper. Psychol.,* 1938, *22*:201-224.

Hovland, C. I., Experimental studies in rote-learning theory: IV. Comparison of reminiscence in serial and paired-associate learning, *J. Exper. Psychol.,* 1939, *24*:466-484.

Hovland, C. I., Experimental studies in rote-learning theory: VI. Comparison of retention following learning to same criterion by massed and distributed practice, *J. Exper. Psychol.,* 1940, *26*:568-587.

Martin, J. R., Reminiscence and Gestalt theory, *Psychol. Monogr.,* 1940, *52,* No. 235.

McGeoch, G. O., The conditions of reminiscence, *Amer. J. Psychol.,* 1935, *47*:65-89.

Ward, L. B., Reminiscence and rote learning, *Psychol. Monogr.,* 1937, *49,* No. 220.

Williams, O., A study of the phenomenon of reminiscence, J. *Exper. Psychol.,* 1926, *9*:368-387.

»»» 17 «««

RETENTION AND FORGETTING — II: QUALITATIVE CHANGES

IN THE course of time, we not only remember *less and less* of what we have learned, but what we do remember frequently suffers serious transformations and distortions. We can increase our insight into the processes of remembering and forgetting if we do not limit ourselves to the measurement of the sheer amounts remembered and forgotten over a period of time. We must also examine carefully the types of qualitative changes and errors which characterize the temporal course of memory. Eventually, such qualitative changes may lend themselves to proper quantification. The twin problems of memory loss (decreases in amount remembered) and memory change (transformations, distortions, and errors) must always be considered together, for they are complementary aspects of the same process.

The Process of Memory Change

The Concept of Trace. In theoretical discussions of memory, it is convenient to refer to the more or less lasting results of learning as *traces.* In its broadest sense, the term *trace* refers simply to the modification which the organism has undergone as a result of learning. Having learned, the organism is no longer the same as it was: learning has left a trace in the organism. Use of the concept of trace does not necessarily imply any assumption about the physiological or neurological counterparts of learning. Trace is a hypothetical concept which helps us develop a convenient terminology about the memory process. One of the tasks of memory theory has been to ask the question: What is the fate of the memory trace in time? How does it develop and change?

The Fate of the Trace in Time. Does the trace merely deteriorate in time, become, as it were, more and more blurred and hence increasingly inefficient in recall? Or does the trace change and develop in systematic ways? A considerable amount of theoretical and experimental argument has centered around this question. Some psychologists have believed that the development of the trace in time follows the same general laws as does perceptual organization (see Chapter 8). According to this view, the trace tends to develop progressively in the direction of greater simplicity, symmetry, and closure so as to conform more and more to the laws of "good" figure. This theory of the "dynamic evolution" of the trace has not remained uncontested by those who believe that the temporal development of the trace is simply in the direction of decreasing accuracy and efficiency. We shall now review the main experimental findings and the methodological problems which arise in attempts to test hypotheses about the temporal development of the memory trace.

The Method of Successive Reproduction

The Nature of the Method. Progressive changes in the nature of the trace must be inferred from progressive changes in retention measured after various retention intervals. For this purpose, many investigators have used the *method of successive reproduction*. In a typical experiment, the stimulus materials (usually geometric designs) are presented to the subject, and he is then required to reproduce them at different time intervals after the original exposure. The important feature of the method is that the *same subject* reproduces the *same materials* on several successive occasions. It was believed that progressive changes in the reproductions would reflect the development of the memory trace of the original stimulus. Geometric designs were generally used as stimulus materials because they are ideally suited for testing the hypothesis that the memory trace develops in accordance with the laws of perceptual configuration. The time intervals over which successive reproductions were obtained have varied and often extended over several months.

Progressive Changes in Successive Reproductions. Consider Fig. 97. The figure shows two stimulus designs and three succes-

sive reproductions of these designs obtained from the *same* subject (a child of school age). The first reproduction was made immediately after presentation of the stimulus, the second drawing was

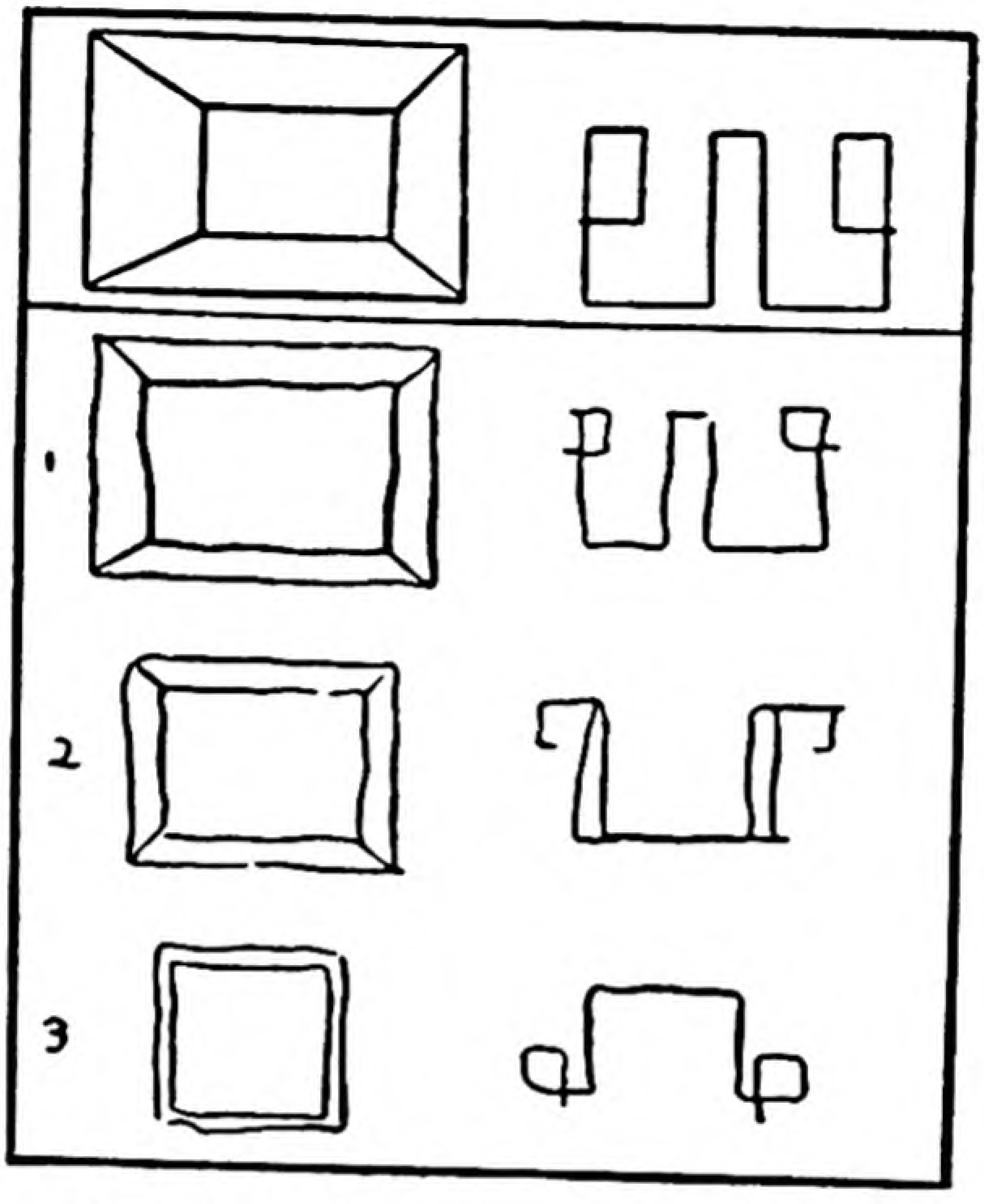


FIG. 97. Progressive Changes in Successive Reproductions of Visual Stimuli. At the top of the figure are the original stimuli and below it three successive reproductions obtained from the same subject (a child): immediately after presentation (*1*), after two weeks (*2*), and after 4 months (*3*). (From G. W. Allport, Change and decay in the visual memory image, *Brit. J. Psychol.*, 1930, *21*:183, by permission of the journal.)

made two weeks later, and the third, four months after exposure to the stimulus material. Clearly, the changes in these successive reproductions are not haphazard but show a progressive trend toward greater symmetry and simplicity. Features of the design which disturb a simple and symmetrical organization, such as the

unequal proportions of the sides in the truncated pyramid and the center stalk in the key, are eliminated. The final product is a simple and compact "good" figure.

It must be noted, however, that progressive unidirectional changes have not always been found in successive reproductions. Frequently, reproductions show progressive changes in some respect but not in others. Again frequently, clear progressive changes fail to appear. When they do occur, progressive changes in reproductions can be described in terms of a few general principles of de-

FIG. 98. This Figure Shows How Differences in Interpretation of a Stimulus Lead to Differences in Reproduction. The labels attached by the subject to the stimulus figure are given below the corresponding reproduction. (From G. W. Allport and L. Postman, *The psychology of rumor*, 1947, p. 140, by permission of Henry Holt and Company, Inc. After J. J. Gibson, Reproductions of visually perceived forms, *J. Exper. Psychol.*, 1929, *12*:1-39, by permission of the journal and the American Psychological Association.)

velopment. First of all, complex figures are *leveled* in the course of time, i.e., many details drop out or are equalized so as to make the resulting figure simpler and more uniform. Complementary to leveling is the process of *sharpening*. Some features of the original figure are not only preserved in successive reproductions but are emphasized and exaggerated. Thus the unequal sides of the truncated pyramid in Fig. 97 are leveled out, but the rectangular frame is preserved and finally sharpened into a double square.

Concurrently with leveling and sharpening, a process of *assimilation* occurs: details of complex stimulus objects are changed so as to conform to normal expectations and established habits of perception. Often such assimilation is due to a verbal label or description

which is attached to a geometrical design. Fig. 98 illustrates the process of assimilation in memory. The figure shows a stimulus and its reproductions by subjects who attached different verbal labels to it. The experimenter did not name the designs, but the subjects did when they first perceived them, and their subsequent reproductions were clearly assimilated to the conventional picture of the object named. Both the initial perception of the object and subsequent memory changes are significantly affected by verbal labels.

The influence of verbal labels on perception and memory can be strikingly demonstrated by intentionally attaching different labels to the same geometrical figure. In one experimental investigation,

FIG. 99. Mnemonic Assimilation to Verbal Labels. Here is a stimulus figure and its reproductions by two subjects to whom it had been presented with different verbal labels—"eyeglasses" and "dumbbell." (From G. W. Allport and L. Postman, *The psychology of rumor*, 1947, p. 143, by permission of Henry Holt and Company, Inc. After L. Carmichael, H. P. Hogan, and A. A. Walter, An experimental study of the effect of language on the reproduction of visually perceived forms, *J. Exper. Psychol.*, 1932, *15*:75, by permission of the journal and the American Psychological Association.)

the subjects were shown a set of geometric designs, each of which resembled two well-known objects. For some subjects, one name was attached to a given design, to other subjects, the alternative name was given. When the subjects later reproduced the designs as accurately as they could, the influence of the verbal labels on their memory was dramatically proved. Fig. 99 shows a stimulus figure which for one subject was called "eyeglasses," for the other, "dumbbell." The reproductions are well assimilated to the habitual appearance of these objects. Since human subjects rely very heavily on verbal responses in most problem and learning situations, assimilation to verbal labels is a critical factor in memory change.

Progressive Changes in Memory for Verbal Materials. The

method of successive reproduction has also been used with verbal materials.[1] The general procedure is the same as in the case of geometric designs. A passage of meaningful material (usually a prose passage) is shown to the subject who is then required to reproduce it at varying time intervals. The progressive changes in the reproductions of verbal materials follow the same general pattern as in the case of visual forms and may be again subsumed under the general headings of leveling, sharpening, and assimilation. As time goes on, the reproduction becomes shorter and shorter, i.e., it is *leveled* due to the omission of an increasing number of items. At the same time, a few features of the original story become dominant (are *sharpened*) and the details of the reproduction are arranged and grouped so as to fit in with the sharpened items. Finally, the reproduction as a whole tends to be reorganized and simplified into a coherent, conventionalized, and easily understandable account (*assimilation*).

Successive reproductions of verbal materials may frequently demonstrate the influence which social attitudes and cultural background exercise on memory change. The process of assimilation often molds and distorts the reproductions in such a way as to conform with the learner's well-established attitudes and cultural expectations. In the course of such assimilation, materials dealing with past events are often modernized, unfamiliar events and circumstances translated into familiar terms, moral and ethical conclusions restated to conform with the subject's own values.

In summary, then, successive reproductions of verbal materials by the same individual may show, like the reproductions of visual forms, progressive leveling, sharpening, and assimilation.

Methodological Criticism of the Method of Successive Reproduction. The method of successive reproduction has been used to gauge the nature of the changes which the memory trace undergoes in time. As a means of discovering progressive changes in the trace, the method of successive reproduction is, however, subject to serious methodological criticisms. First of all, a subject's reproduction cannot be simply regarded as an index of the state of

[1] Notably by F. C. Bartlett, whose book, *Remembering* (Cambridge: Cambridge University Press, 1932), describes a monumental series of experiments dealing with the problem of memory change.

the trace. In reproducing a geometrical design, for example, the subject is limited by his drawing ability. He may fail to put down on paper everything he remembers as he remembers it. A more serious difficulty, however, stems from the fact that the very act of reproduction probably changes the nature of the trace. Thus, in making *successive* reproductions, the subject is influenced not only by his original perception and the trace formed by it but also by his preceding reproductions. If he made an error or introduced a distortion during the first in a series of recalls, he remembers his error when he attempts his second reproduction, and so on. As Woodworth has pointed out, "We cannot observe the state of a memory trace without letting it act and perhaps distorting it and altering its subsequent history."[2] To summarize, attempts have been made to follow the temporal development of the memory trace by comparing successive reproductions of the same material by the same individual, but the successive acts of reproduction themselves seriously affect the nature of the trace.

Comparison of Successive and Single Recalls

The Method of Single Reproduction. To meet, at least in part, the methodological objection to the method of successive reproduction, the *method of single reproduction* has been used. Under this procedure, equated groups of different subjects reproduce the same stimulus materials but at different time intervals after original learning. Reproductions obtained at different time intervals are then compared with each other. There are striking differences between the pictures of forgetting yielded by the methods of successive and single reproduction. First of all, the two methods result in completely different amounts of forgetting. As Fig. 100 shows, the method of single reproduction shows a steadily increasing loss in time—a more or less typical forgetting curve. With successive reproductions, on the other hand, the changes in amount retained are much slower. Each successive reproduction serves as a rehearsal or learning trial, serves to fixate and perpetuate what has been initially retained.

The *qualitative* differences between the results obtained with the

[2] R. S. Woodworth, *Experimental psychology*, New York: Henry Holt and Co., 1938, p. 91.

two methods are equally striking. Whereas the method of successive reproduction often yields progressive changes in time, no such trends have been found when the single reproductions made by different subjects after varying time intervals are compared. In the case of single reproductions, it has been found that recall becomes less and less accurate as the interval between original learning and

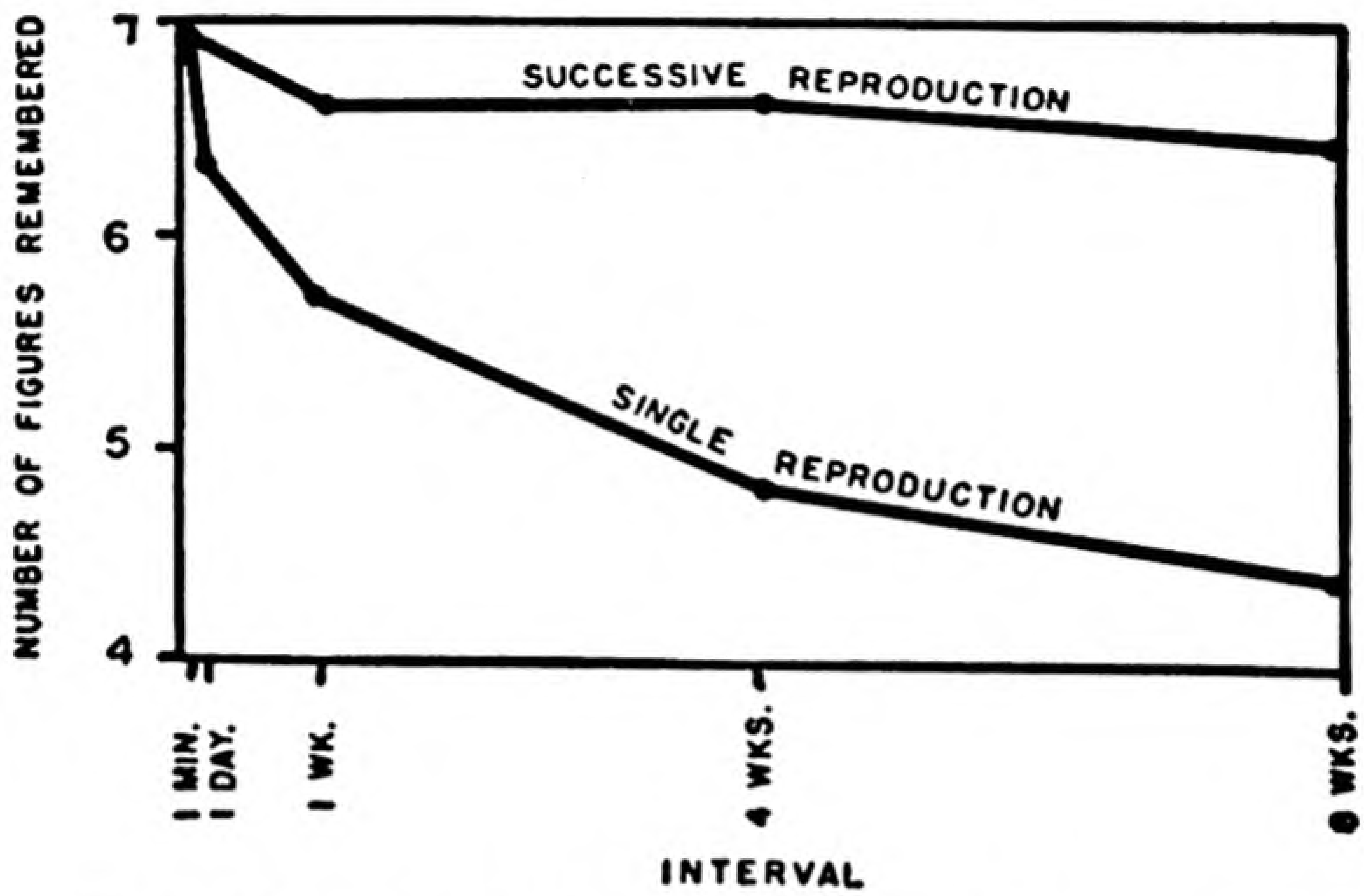


FIG. 100. Comparison of Retention Scores Obtained by the Method of Successive Reproduction and the Method of Single Reproduction. (From J. A. McGeoch, *The psychology of human learning,* 1942, p. 360, by permission of Longmans, Green & Co., Inc. After N. G. Hanawalt, Memory trace for figures in recall and recognition, *Arch. Psychol.,* 1937, *31,* No. 16, p. 25, by permission of the journal and the American Psychological Association.)

retention test is lengthened but there is no evidence for progressive changes. There is, rather, what has been described as "true forgetting"—fewer and fewer items are remembered and what is remembered becomes less and less accurate.

Similar results have been obtained with the *method of single recognition.* Equated groups of subjects are given recognition tests for the original materials at varying time intervals. There is no evidence for progressive changes toward simplicity, symmetry, and closure in the figures recognized by the subjects. Again, only decreasing accuracy of recognition behavior is found as a function of

time. *Successive recognitions* by the same subjects, on the other hand, have in some cases shown progressive changes in the direction of "better" forms.

The radically different pictures of forgetting yielded by the methods of single reproduction (recognition) and successive reproduction (recognition) underline the methodological dilemma which we face when we attempt to chart the temporal development of the memory trace. In successive recalls by the same subject, each test is profoundly influenced by the preceding tests. Recalls by different subjects at different time intervals obviate this difficulty, but they do not yield information about the temporal development of any one single system of memory traces, for individuals, however well equated, are not interchangeable.

Methodological Outlook. Our discussion leads to the conclusion that there is no foolproof method for studying the temporal development of one individual's memory traces. Nevertheless, the types of progressive changes which are obtained by the method of successive reproduction are worth studying because of the light which they help throw on the kinds of memory functions with which we have to deal continually in practical situations. True, each successive reproduction is influenced by the preceding ones, but is this not exactly what happens in daily remembering? If we want to recall a childhood experience, we cannot tap the pure trace of that event, but we must perforce recall it, warped as it is by all its previous reproductions. To identify and study the changes which arise in the course of successive recalls of the same event remains an important task for the experimental psychology of memory.

The Method of Serial Reproduction

The Nature of the Method. The omissions, changes, and distortions which characterize successive reproductions by the same individual are exaggerated and accelerated in *serial reproductions.* Under this procedure, only the first of a group of subjects is exposed to the stimulus materials. The "eyewitness" passes on what he remembers to the second subject, who in turn transmits it to a third, and so on, until a chain of reproductions has been completed. The initial stimulus material may be a prose passage, a picture, or any other item exceeding the span of immediate memory. It is advisable

to employ stimulus materials which are rather rich in details, thus providing an opportunity for various types of memory change to unfold themselves in the course of serial transmission. The presentation of the material may be either visual or auditory. The time interval between presentation and transmission of the material may

FIG. 101. A Stimulus Picture Which Has Been Employed in a Series of Experiments on Serial Reproduction. A typical terminal report, the last in a chain of six follows: "Picture of a trolley car with seven people. There is a woman with a baby. There are some colored people. Someone is flashing a razor blade." (From G. W. Allport and L. Postman, *The psychology of rumor*, 1947, p. 71, by permission of Henry Holt and Company, Inc.)

be varied. Good results are obtained even if the successive links in the chain follow each other immediately.

Memory Loss and Change in Serial Reproductions. A series of reproductions by a chain of individuals yields the same general kinds of memory loss, change, and distortion as successive reproductions by the same individuals, but the speed and the magnitude

of the changes are greatly enhanced. The nature and extent of sucn changes are most easily illustrated with the aid of a concrete example.

Fig. 101 shows a picture which has been employed in a series of experiments on serial reproductions. This picture was especially designed for use in such experiments. It is rather dramatic in quality and contains a large number of details which are easily subject to error and distortion. The picture was projected on a screen and shown to the first subject who described it in considerable detail to the second subject (the latter, of course, was unable to see the screen himself), the second subject transmitted what he had heard to the third subject, and so on.

Under Fig. 101 we present a typical terminal report, the last in a chain of six. The initial description was accurate and rich in details. By the time it has traveled down to the final link in the chain, the description is not only inaccurate but also exceedingly short—a highly condensed, skeletonized account. In the course of serial transmission, the story has been rapidly *leveled.* Considerable leveling invariably characterizes serial reproductions until the accounts become short, concise, easily told and remembered. As Fig. 102 shows, leveling takes a continuous progressive course during a series of reproductions. The initial rate of leveling is quite steep. Early in the series, each subject loses a considerable proportion of what is presented to him. Later on, however, the reproductions become so short that they may be repeated virtually by rote. Thus leveling, though extensive, never leads to complete obliteration.

The speed with which leveling proceeds under conditions of serial reproduction is noteworthy. In the experiments in which Fig. 102 was obtained, successive subjects transmitted the reproductions immediately after hearing them, and yet the loss of detail is extremely rapid. With serial transmission, there is as much leveling in a few minutes as there may be after weeks with single reproductions. Selective perception and retention of the material occur not once but as many times as there are subjects in the series. The effects cumulate and a mere skeleton of the original material remains at the end of the series.

Where there is leveling, there must necessarily be *sharpening,* for sharpening is the selective emphasis on a few features or details

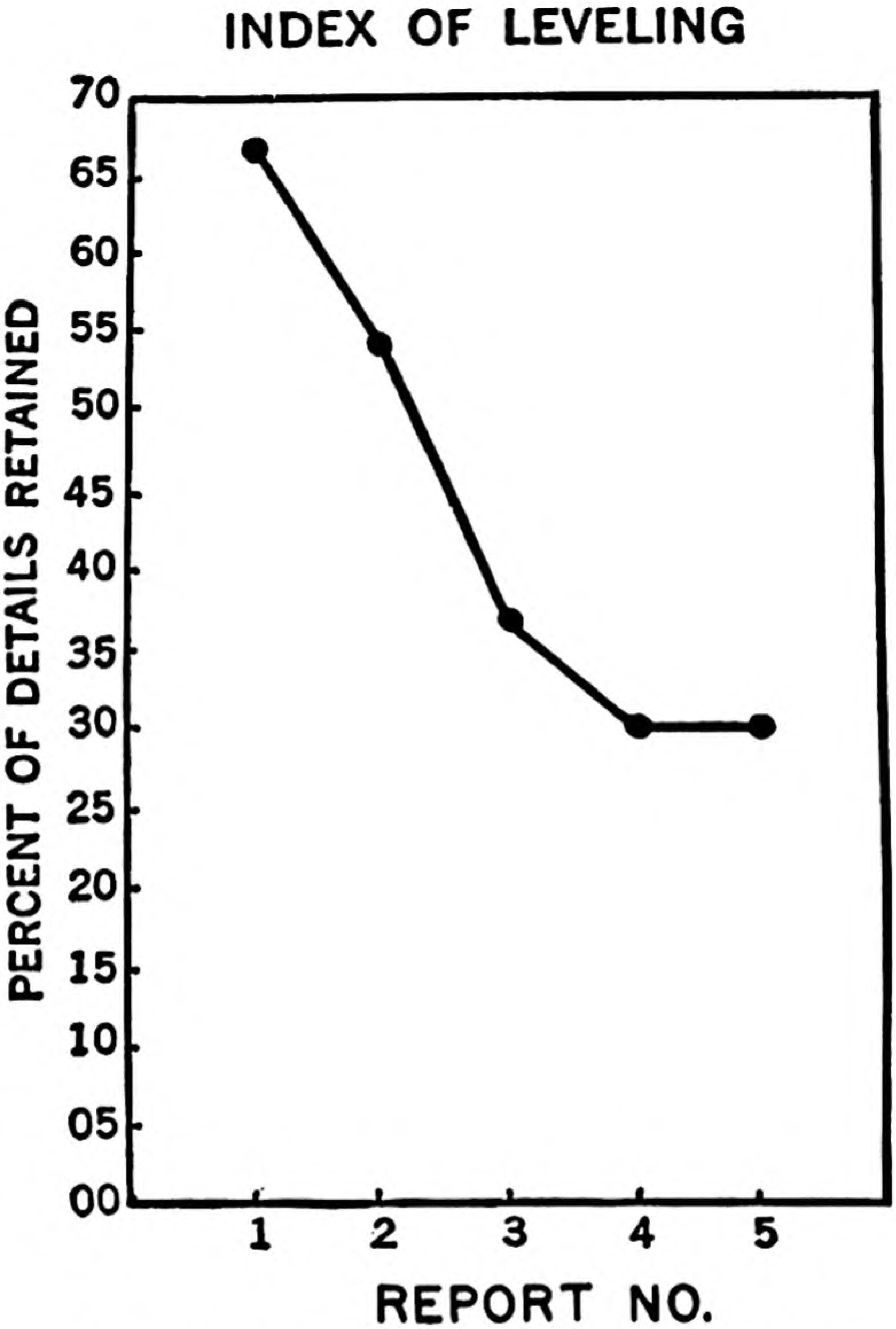


FIG. 102. The Course of Leveling in Serial Reproduction. The curve shows the percent of the original reported details retained in a series of reproductions. (From G. W. Allport and L. Postman, *The psychology of rumor*, 1947, p. 76, by permission of Henry Holt and Company, Inc.)

from a larger context. As the terminal report under Fig. 101 illustrates, a few of the many varied details of the picture were retained until the end of the series. By the very fact of their selective reten-

tion at the expense of other items, they may be considered sharpened. Frequently, moreover, such selected items are exaggerated and magnified out of all proportion and come to dominate the reproductions. Sharpening is by no means a haphazard matter. If the material is visual, certain physical characteristics such as relative size, movement, and brightness may determine what is sharpened. These are the well-known determinants of "attention." The presence of familiar objects and symbols, verbal labels attached to parts of the material, may result in selective sharpening. Last but not least, interests and attitudes are ever-present, important determinants of selective perception and retention. Out of a multitude of items, we select and remember those which fit in with our interests and values, or which conform to our habits and expectations.

Finally, in serial as in successive reproductions, *assimilative* processes are constantly at work. Details are grouped around a principal theme so as to make the reproductions more coherent and meaningful. That which is unfamiliar and unusual is rendered familiar and conventional. Facts may be distorted in order to be assimilated to the individuals' attitudes and values. In summary, then, the major changes in serial as well as successive reproductions can be described in terms of the three-pronged process of leveling, sharpening, and assimilation.

In serial reproductions, the speed and magnitude of memory distortions are at a maximum. When successive reproductions are made by the same individual, the distortions are held in check, at least to a certain extent, by the fact that he had been exposed to the original stimulus materials and was able to identify and label them. In serial reproductions, no such checks can operate. Each successive subject in the series has nothing to rely on except a report (frequently confused and halting) from another individual. His retention is not anchored to any stable points of reference; a subject who is merely a link in a chain has no "general idea" of the nature or content of the original stimulus materials to which he can cling in his attempts at recall. The extent to which serial reproductions can stray away from the initial stimulus is dramatically illustrated in Fig. 103. We see a schematic drawing of an owl, passed on from subject to subject, gradually transformed into a fanciful drawing of a cat. Somewhere along the line the critical mistake was made

and the verbal label "cat" instead of "owl" was attached to an ambiguous drawing. This mistake was irreversible and thenceforth the errors cumulated as the reproductions were assimilated more and more to the new conception. It is more than unlikely that distortions would ever follow a similar course in successive reproductions by

FIG. 103. How an Owl Became a Cat in a Series of Reproductions by Different Subjects. (From F. C. Bartlett, *Remembering*, 1932, pp. 180-181, by permission of the Cambridge University Press.)

the same individual. Herein lies the critical difference between successive and serial reproductions. Successive reproductions by the same individual, no matter how distorted, are likely to develop within definite limits set by the original perception. The picture of the owl might change in many ways, but an owl it would remain once it had been identified. In the absence of such constraints, there is virtually no limit to the cumulation of error and distortion in serial reproductions.

The Continuity of Perception, Memory, and Report

The analysis of memory loss and change points up the basic continuity of perceiving, remembering, and reporting. These three activities are inextricably fused and jointly determine measurable retention performance.

The process of memory is launched on its course by the learner's initial perception of the stimulus situation. Perception is selective, and out of the totality of stimuli present only a limited fraction is perceived. Only those events which are favored by selective perception are well retained. Not only is perception selective, but it entails active organization and interpretation (e.g., with the aid of verbal labels) of the stimulus materials. Memories of past events and experiences, learned habits of seeing and responding, profoundly influence the initial perception and hence subsequent memory. Under the directive influence of the initial perception, changes continue and cumulate—leveling, sharpening, and assimilation take their course. When the time has come for active recall, the individual attempts to *reconstruct* his past experience, and in the process of reconstruction the continuous series of omissions, changes, interpretations, and distortions which began at the very first moment of perception finds its full expression. The act of recall itself, the ability to reproduce or report what one remembers, is a final source of memory change. Memory, then, is a constructive, creative process, a process which actively unfolds itself from the very first moment of initial perception. No analogy could be less apt than the comparison of memory to an image fixed on a photographic plate, an image which fades in time but is available for reproduction as long as its outlines have not hopelessly blurred. On the contrary, whenever an individual remembers, he re-creates his past experience, subject to all the errors and transformations which have accumulated since he first perceived the event which he is trying to remember.

Memory Changes and Testimony

The Observer as Reporter. Nowhere are the continuity and interdependence of perceiving, remembering, and reporting demonstrated more fully than in the psychology of testimony. The psychology of testimony has been aptly described as the study of the

"observer as reporter." Testimony was one of the first fields of practical application of experimental psychology, precisely because in its study the newly gained knowledge of the processes of perception and memory could be used to practical advantage.

The psychology of testimony is concerned with two broad classes of problems: (1) How much does an eyewitness perceive of an event at which he is present, and how accurately and clearly does he perceive it? (2) How accurate and reliable is his report of the event after a period of time? The practical implications of the answers to these questions for such fields as trial law and newspaper reporting are self-evident.

Types of Testimony Experiments. Laboratory experiments on testimony fall into two main classes: (1) picture tests, and (2) reality experiments. In a picture-test situation, a complex scene is shown to the subject, and he is later required to remember as many details as he can. In a reality experiment, some dramatic incident is unexpectedly enacted in the presence of a group of subjects. An unexpected intrusion into the classroom, a well-rehearsed, violent quarrel, threatened physical attack on the professor are favorite classroom situations for reality experiments. Afterward, the eyewitness' memory for the details of the incident is tested.

Test Methods in Testimony Experiments. As in all memory experiments, the results obtained in testimony experiments vary with the methods of measurement employed. Two basic test methods have been used: (1) free narrative, and (2) interrogatory or cross-examination. In the method of free narrative, the subject is simply requested to give as full an account of the scene or incident as he can. He is allowed to do so without any questions or guidance from the experimenter. For an interrogatory, a long list of questions is prepared which the subject must answer. These questions carefully probe the witness' memory for all the details of the situation. The interrogatory is likely to give a much more complete picture of the witness' memory for the event, since the questions serve to remind him of details which he may otherwise have failed to mention. This advantage of the interrogatory is, however, a double-edged one. Subjected to an interrogatory, the witness may easily fall victim to suggestion and give the answers which he believes to be implied in the questions. Probably a judicious combination of the two meth-

ods—a spontaneous account followed by carefully worded questions—will yield the most sensitive measures in testimony experiments.

The Fallibility of Witness. Testimony experiments provide dramatic demonstrations of the inaccuracy of human observers and the fallibility of their memory. Experimental results almost invariably show that the witness' perception and retention of the situation are highly selective and limited. He fails to notice or quickly forgets many of the details, but he usually remembers those which help to explain and interpret the situation as he conceives of it. Most of the other items he fails to remember. Names, places, and times are especially subject to confusion and forgetting. In most cases, the result is a highly elliptical and subjectively colored account.

Anything that is unusual or unfamiliar has a small chance of being correctly observed and remembered by a witness. The experimental records show again and again that events which are out of the ordinary are either assimilated in memory, i.e., recast in terms of normal expectations, or exaggerated out of proportion and made the central theme of the report. This inability to observe and report the unusual correctly is especially pronounced under conditions of excitement and emotional stress. And yet, it is precisely in relation to unusual occurrences that the testimony of eyewitnesses is most important.

The experimental investigations of testimony have served to underline the fallibility of observation and report, but the findings regarding the inaccuracies of testimony can be easily exaggerated. The experimental records show a considerable amount of correct recognition and recall. There are substantial positive correlations between age and intelligence on the one hand, and fidelity of testimony on the other. If the time interval between observation and report is not too long, much of the testimony of intelligent adults is often fairly reliable and accurate.

The official testimony of an eyewitness may not be his first report of the incident at which he was present; he may have previously told it to others. In effect, the witness may give successive reproductions of the same event. The nature and limitations of successive reproductions must, therefore, be borne in mind in evaluating such testimony. In giving consecutive reports, what the individual recalls is

likely to be considerably influenced not only by his memory of the event proper but also by his memory of his previous reports. Initial distortions and errors may be fixated and perpetuated in the course of successive reports. The sooner after the event the critical or "official" testimony is obtained, the less danger there is of perpetuation of errors by successive reproductions.

Memory Change and Rumor

Rumor and the Method of Serial Reproduction. Memory change plays a major role in another important area of social communication: *rumor.* A rumor may be defined as a communication which is passed on from individual to individual, unsupported by reliable evidence. Throughout history, rumors have been a favorite unofficial source of "news," rife especially in times of emergency and crisis.

The method of serial reproduction provides the experimental paradigm for rumor. As in the case of many rumors, this type of experiment starts with a fact or event perceived by an eyewitness and passed on by word of mouth from individual to individual. The types of memory change and loss revealed by the method of serial reproduction are, indeed, characteristic of many rumors. As a rumor travels from person to person, it tends to become shorter, crisper, acquiring a slogan-like quality (*leveling*). Rumors are characterized by selective emphasis and frequent exaggeration of some features of an event at the expense of others (*sharpening*). Finally, in rumor, the process of *assimilation*—the powerful influence on memory change of habits, interests, and values—comes fully into its own. Rumors, then, may be regarded as the results of serial reproductions in social life.

Laboratory Study of Rumor. It is often difficult, if not impossible, to follow the course of a rumor in real life, to record the changes and distortions as a story is passed on from individual to individual. The experimental study of rumor, therefore, relies primarily on the method of serial reproduction. A laboratory rumor simply consists of the telling and retelling of a story by a series of subjects.

Although this method brings us as close as we can ever come under the controlled conditions of the laboratory to the study of

rumor, important differences between laboratory-created and real-life rumors remain.

1. Subjects in a serial reproduction experiment will have a set toward maximum accuracy of reproduction by virtue of the very fact that the investigation is conducted in the laboratory. Rumor merchants in daily life rarely have such concern about the accuracy of the information which they pass on.
2. The practical exigencies of experimentation will usually make it necessary to use only short time intervals (often of the order of minutes or hours) between successive links in the rumor chain. In daily life, the times which elapse between the hearing and telling of a rumor are varied and frequently much longer.
3. The personal relationship between the members of the chain is likely to be very different in a laboratory situation from what it is in the transmission of actual rumors. Rumors usually travel among individuals with common interests and motives. Accounts are often passed on as part of intimate personal conversation, in situations altogether different from a laboratory setting for an experiment on serial reproduction.
4. Finally, and perhaps most important of all, the subjects in a serial-reproduction experiment and the rumor monger differ radically in their motivation. The laboratory subject tries to carry out, to the best of his ability, the experimenter's instructions. In passing on the report which is presented to him, he does not try to grind a personal ax. How different the motivational picture is in the case of the rumor monger! The tales which he receives and transmits are often actuated by deep-seated fears, inspired by hates and aversions of long standing, embody the fulfillment of hopes and wishes. To these dominant motives, the stories he spreads are inevitably assimilated. Unlike the laboratory subject, the rumor agent often has a deep personal stake in the story which he tells.

We have examined the differences between a "laboratory rumor" and the type of rumors which we so frequently encounter and have to fight in daily life to illustrate the type of problem which experimental work of social psychology so frequently faces. The experimental social psychologist must begin with a laboratory paradigm, such as the serial reproduction model for the study of rumor. In

many respects, the paradigm will fail to reproduce the conditions under which the social event occurs. But the experimental paradigm is nevertheless very much worth while: from the experimental paradigm we obtain the variables, concepts, and hypotheses in terms of which we may proceed to plan the investigation of the "real thing" in the field. Armed with knowledge of the nature and determinants of memory change, we may be better able to deal with the practical problems of testimony and rumor. The laboratory work helps to define the variables and to provide the essential methodological tools for the researcher in the field.

Experiment XXIV
Memory Change in Serial Reproduction

Purpose. To investigate the amount of memory loss and the nature of memory change under conditions of serial reproduction.

Materials. Both verbal and visual stimulus materials may be used. Whether verbal or visual, the stimulus material should be complex and contain a number of details well beyond the subject's immediate span of memory. A scene such as the one shown in Fig. 101 can be used with good results. We shall describe the procedure for an experiment using such a visual stimulus. The procedure, however, remains substantially unchanged if verbal material is used.

Apparatus. It is most convenient to present the stimulus picture by projecting it on a screen. A lantern slide of the picture should be prepared and then exposed by means of a standard projector. The screen should be placed so that successive subjects can enter the room without seeing it.

Procedure. Six or seven subjects are sent out of the room. The slide is projected on the screen throughout the demonstration. The first subject is called in and he is requested to stand in a position from which he cannot see the screen. The experimenter (or some other person) then gives him a detailed description of the scene (including about twenty to twenty-five items), preceded by the following instructions:

> "There is a picture on the screen which I am going to describe to you in some detail. I want you to listen carefully. You will be asked to repeat what you have heard to the next subject."

After the initial "eyewitness" description has been given, the second subject is called in and asked to take his place beside the first. He is given the following instructions:

"Mr. (Miss) X will transmit to you a description of the picture on

the screen which he has just heard. I want you to listen carefully. You will be asked to repeat what you have heard to the next subject."

(These instructions are used for all other subjects in the series.)

The first subject then gives his reproduction of what he has heard and takes his seat. The third subject is called in, and the procedure is continued until the entire series of reproductions has been obtained.

Records. It is important to obtain as accurate a record as possible of each reproduction, including the initial one. Ideally, a device, such as a wire recorder, should be used. If an automatic recording apparatus is not available, a quick note-taker usually can obtain a complete record, especially if the subjects are instructed to speak slowly and distinctly. Of course, the reproductions may be obtained in written rather than oral form. In that case, however, the experiment cannot be used as a classroom demonstration.

Audience. The presence of an audience will have an important effect on the subjects' behavior. Conscious of being under observation, the subjects are likely to be very cautious and careful, being afraid to "make fools of themselves." They will tend to repeat only items of which they feel sure. Reports obtained in the presence of an audience will, therefore, be shorter and more conservative than those obtained without an audience. An interesting variation of the experiment is to obtain reports with and without an audience and to compare them.

Treatment of Results. The scoring of the records will necessarily involve a certain amount of subjective interpretation. It is advisable, therefore, to have several experimenters score each record independently. Disagreements and differences of interpretation should be discussed until a consensus is achieved.

The records may profitably be scored with a view to the three dimensions of memory change: leveling, sharpening, and assimilation.

1. *Leveling.* For each reproduction, the number of details correctly remembered is scored. Under the conditions of this experiment, an item need not be remembered verbatim in order to be scored as correct; it must, however, be retained without substantial alteration of meaning. The number or percentage of correctly retained items is then plotted as in Fig. 102, to depict the temporal trend of leveling. The significance of the differences between successive points can be tested statistically.
2. *Sharpening.* In evaluating the reproductions, special attention should be paid to the items which are retained throughout the series of reproductions (sharpened). What are the characteristics of these items? Were they outstanding in size, color, degree of movement? Were

they familiar or unfamiliar items? Were the sharpened items exaggerated or magnified? Did the sharpened items come to dominate the reproductions, resulting in a shift in the entire theme of the description?

3. *Assimilation.* In looking for evidence of assimilation, search the reproductions for errors and distortions which can reasonably be ascribed to the influence of established habits, expectations, and attitudes. If an unfamiliar feature is changed so as to conform with what is usual and expected, if an odd phrase is turned into a well-worn cliché, you have an example of assimilation. If the distortions and errors seem to reflect the teller's wishes, fears, or antipathies, they are probably examples of assimilation. If there has been a shift in the general theme of the reproductions, show how specific details are changed so as to fit in with the new theme. This type of analysis is somewhat subjective and rather difficult at first, but it is a rewarding exercise in the interpretation of qualitative data.

General References

Bartlett, F. C., *Remembering*, Cambridge: Cambridge University Press, 1932.

Katona, G., *Organizing and memorizing: studies in the psychology of learning and teaching*, New York: Columbia University Press, 1940.

Koffka, K., *Principles of Gestalt psychology*, New York: Harcourt, Brace and Co., 1935, Chap. 11.

McGeoch, J. A., *The psychology of human learning*, New York: Longmans, Green and Co., 1942, Chap. 8.

Woodworth, R. S., *Experimental psychology*, New York: Henry Holt and Co., 1938, Chap. 4.

Method of Successive and Single Recall

Allport, G. W., Change and decay in the visual memory image, *Brit. J. Psychol.*, 1930, *21*:133-148.

Carmichael, L., Hogan, H. P., and Walter, A. A., An experimental study of the effect of language on the reproduction of visually perceived form, *J. Exper. Psychol.*, 1932, *15*:73-86.

Clark, K. B., Some factors influencing the remembering of prose material, *Arch. Psychol.*, 1940, No. 253.

Gibson, J. J., The reproduction of visually perceived forms, *J. Exper. Psychol.*, 1929, *12*:1-39.

Hanawalt, N. G., Memory trace for figures in recall and recognition, *Arch. Psychol.*, 1937, No. 216.

Hebb, D. O., and Foord, E. N., Errors of visual recognition and the nature of the trace, *J. Exper Psychol.*, 1945, *35*:335-348.

Serial Reproduction, Testimony, and Rumor

Allport, G. W., and Postman, L., *The psychology of rumor*, New York; Henry Holt and Co., 1947.

Kirkpatrick, C., A tentative study in experimental social psychology, *Amer. J. Soc.*, 1932, *38*:194-206.

Stern, W., *General psychology from the personalistic standpoint*, New York: The Macmillan Company, 1938, Chap. 14.

Whipple, G. M., The observer as reporter: a survey of the "psychology of testimony," *Psychol. Bull.*, 1909, *6*:153-170.

»»» 18 «««

TRANSFER OF TRAINING

THE effects of learning and conditioning are cumulative in time. When an organism learns a list of words or undergoes conditioning, it is more or less permanently modified by this experience. When this organism faces new tasks and new problems, its behavior may be seriously affected by the results of past learning and conditioning. It is only through such cumulative effects of learning that steady intellectual development and growth, progressive refinement of skills, and creative thinking are made possible. Sometimes, it is true, past experience hinders rather than aids the acquisition of new skills, but the balance is overwhelmingly on the credit side. The effect of past learning on new learning is designated as *transfer of training*. Transfer of training is one of the most pervasive characteristics of behavior, for it is transfer of training which guarantees the continuity and lawful development of habits of ever-growing complexity. The emergence of consistent personality traits no less than the organization and integration of knowledge depends on transfer of training.

Types of Transfer

The effects of past learning on new learning may be classified in one of three categories:

Positive transfer effects occur if experience facilitates the acquisition of a new skill or the solution of a new problem. Placed in the new situation, the learner performs significantly better than he would without the benefit of past training.

Negative transfer effects are inferred if past experience renders more difficult or slows down the acquisition of a new skill or the solution of a new problem. Placed in the new situation, the learner performs more poorly than he would without previous training.

Zero transfer effect denotes the fact that performance in the new situation is neither aided nor hindered by past training. A statement that there is zero transfer can mean only that with the measuring devices at our disposal no transfer effects from one situation to the other can be detected. Where one method of measurement shows zero transfer effect, another more sensitive one may reveal either positive or negative effects.

Design of Transfer Experiments

Use of Equated Groups. One basic experimental paradigm calls for the use of an experimental group and a control group. Suppose we wish to investigate the transfer effects from memorizing nonsense syllables to memorizing prose. We select a group of subjects and set them the task of memorizing a prose passage (Task T_1). On the basis of their performance, we divide our subjects into two groups matched for ability to memorize prose. One of these groups serves as the experimental group, the other as the control group. The experimental group is given practice in memorizing nonsense syllables (Task T_2), while the control group rests or engages in some activity which is quite unrelated to memorizing. Finally, both groups are given a new test on memorizing prose material. The *difference* in the gain (or loss) of the two groups from initial test to final test is a measure of the transfer effect. Thus if the experimental group shows a greater improvement on the final test than does the control group, we may attribute this difference to the positive transfer from the interpolated practice. The general design of this experiment may be summarized as follows:[1]

Experimental Group:
Initial Test on T_1——Training on T_2——Final Test on T_1
Control Group:
Initial Test on T_1——Rest——Final Test on T_1

This design allows an experimental isolation of the factor of trans-

[1] The reader should compare this design with that of the retroactive inhibition experiment on p. 370. The formal structure of the two designs is the same. In the case of the transfer experiment, we measure the effect of past learning on new learning. In the case of the retroaction experiment, we are concerned with the effects of new learning on retention of old material. Retroactive inhibition may thus be considered an example of negative transfer effect.

fer. Improvement on the final test may be due to factors other than interpolated practice, such as the experience provided by the initial test. By treating the two groups alike in all respects, except for the interpolated training period, it is possible to ascribe differences in gains or losses to this training. Clearly, this experiment can yield valid results only if the two groups are equated with respect to their performance on the initial test. Benefit from training varies with initial ability. Typically, a poor memorizer can derive considerable benefit from practice—there is so much room for improvement. A skillful memorizer, on the other hand, cannot improve much more by practice. It is only by equating the initial performance level that we render the gains or losses of the two groups comparable.

Instead of using a foretest, it may be possible to equate the two groups by means of an outside criterion. For example, if the transfer value of different methods of teaching is to be evaluated, subjects may be equated on the basis of examination grades. The foretest may then be omitted and the design shortened as follows:

Experimental Group:	Training on T_2——	Test on T_1
Control Group:		Test on T_1

On the assumption that the two groups were initially equal, any difference on the test of T_1 may be ascribed to the effect of training. This procedure has the virtue of being brief, but it has little else to recommend it. It is always safer to equate the two groups explicitly on the activity to which transfer is expected.

Use of Equated Tasks. Instead of equating subjects, we may equate tasks and use only one group of subjects. Suppose we wish to study the transfer from solving one set of mathematical problems, T_1, to solving another set of problems, T_2. T_1 and T_2 are two standardized sets of problems known to be of equal difficulty. A group of subjects solves the two sets in succession. If their performance on the second set is superior to their performance on the first, we may attribute this gain to positive transfer. This interpretation would, of course, depend on the correctness of the assumption that the two sets of problems are strictly equal in difficulty. Such an assumption should not be lightly made in the absence of thoroughly reliable norms of difficulty.

A Counterbalanced Design. Another experimental procedure again employs two groups but instead of dividing them into an experimental and a control group, it uses a *counterbalanced* design. Suppose we wish to measure the transfer from mastering one maze to the learning of a second maze. We have, of course, to use two mazes, T_1 and T_2. We attempt to make these two mazes approximately equal in difficulty although it may not be possible to make them exactly so. The members of one group first learn maze T_1 and then maze T_2; the other group reverses the order, T_1 is learned after T_2. We then compare the average performance of the two groups on the first task with their average performance on the second task, using time scores, error scores, or whatever measure of learning is most appropriate. The design of the experiment may be summarized as follows:

Group I: T_1 followed by T_2
Group II: T_2 followed by T_1

If the average performance on the task learned second is significantly better than on the task learned first, there has been positive transfer. If there is no such significant difference, we conclude that there has been zero transfer. Finally, if performance on the second task is significantly poorer than on the first, there has been negative transfer.

The design outlined above has the advantage of making unnecessary the strict equation of either subjects' ability or difficulty of tasks. Both groups of subjects practice both tasks in counterbalanced order. Still, if the two groups differ greatly in ability or the two tasks in difficulty, disturbing interaction effects may occur. If an able group has to go from an easy to a hard task and a poor group from a hard to an easy task, it would be difficult to compare average performance on the first task with average performance on the second task. Minor variations in ability and difficulty, on the other hand, are adequately controlled by this design.

Degree of Learning. As in all learning studies, degree of mastery is an important variable in the transfer experiment. The activities from which and to which there is transfer can be practiced to varying extents, and the conclusions concerning transfer may be seriously affected by these variables.

Consider first the activity *from which* there is transfer. Not only the amount but also the sign of the transfer effects (positive or negative) may depend on the amount of practice in this activity. If the transfer effects from a first to a second task tend to be positive, the amount of transfer will vary with the degree to which the first task has been mastered. For example, if the first of two mazes has been poorly learned, the learning of the second maze will be little facilitated. Thorough mastery of the first maze, on the other hand, may greatly accelerate the learning of the second. The evidence strongly points to the conclusion that the beneficial effect of past training on new learning is positively correlated with the amount of that training.

Negative transfer effects from a first task to a second also show considerable dependence on the degree to which the first task has been learned. Negative transfer effects are likely to appear when the first task has not received extensive practice. With increased practice in the first task, negative transfer effects may turn into positive.

The degree to which the second task is learned is also important. In some experiments, there is only one trial of the second activity and amount of transfer is computed on the basis of this test trial. For example, in gauging transfer from one maze to another, the first maze may be learned to the criterion of one errorless run, and then one test trial given on the second maze. It is possible, and often desirable, however, to carry the second activity to the same criterion as the first, e.g., to practice the second maze until an errorless run is achieved. In that case, transfer effects at successive stages of the second learning activity can be measured, and additional measures of transfer, such as amount of time or number of trials saved, can be obtained. Experiments using this procedure have tended to show that the main transfer effects appear early in the learning, although the experimental (transfer) group may retain a small advantage over the control (nontransfer) group over the total learning period. By carrying out the second activity to a criterion of mastery, we not only can study the temporal course of the transfer effects but also provide adequate opportunity for these effects to manifest themselves fully. In situations in which an initial readjustment to

the second situation is necessary, use of only one test trial may indeed prevent transfer effects from appearing or minimize them.

In any transfer experiment, then, the degree to which each of the activities is learned is an important parameter. It is often necessary to vary this parameter in order to obtain a full picture of the transfer effect.

Interval Between the Two Learning Activities. In planning an experiment on transfer of training, another variable which deserves careful attention is the time interval between the learning tasks. We must consider (1) the nature of the activities filling the interval, and (2) the length of the interval.

If we are interested in measuring transfer from a first task, T_1, to a second task, T_2, it is important that no uncontrolled activities intervene between T_1 and T_2 which could obscure the transfer effects. If an activity, T_3, similar to T_1, were to occur during this interval, it might either increase or decrease the transfer effects under consideration. Thus if we are measuring the transfer effects from one memory task to another, we must make sure that there is no uncontrolled memory practice between the two experimental tasks. For this reason, it is advisable to space the experimental tasks so that the activity during the interval between T_1 and T_2 can be controlled by the experimenter. Of course, it is never possible to control the activity of a subject completely, but by engaging him in some work quite unrelated to the critical tasks, it is possible to minimize uncontrolled interference.

As for the time interval between T_1 and T_2, there is some experimental evidence that transfer effects remain virtually constant over a long period of time. In one study, for example, transfer effects from T_1 to T_2 (ideational learning problems) were measured after intervals ranging from two days to ninety days. Lengthening of the time interval did not produce any significant decreases in transfer effect. This result is at first surprising since degree of transfer is a measure of retention, and retention typically shows a steady decline in time. It appears that the type of behavior change which constitutes transfer is resistant to forgetting, probably because transfer effects frequently do not depend on the retention of specific responses but rather on the acquisition of useful methods and approaches to new problems.

WHAT IS TRANSFERRED IN TRANSFER OF TRAINING

The Search for Elements. One theory of transfer is the so-called theory of *identical elements.* Since sign and amount of transfer effects vary with specific materials and conditions of training, this theory held that amount of transfer is proportional to the number of elements common to the activities between which transfer occurs. If we think along the lines of this theory, we must analyze T_1 and T_2—the activity from which there is transfer and the activity to which there is transfer—and break them down into constituent elements, say, stimulus and response elements. We should, then, expect transfer to increase as the number of elements common to T_1 and T_2 increases.

One of the early experiments carried out by psychologists working in the framework of this theory will illustrate this approach. Subjects were trained in estimating the areas of rectangular designs within a limited size range. The transfer from this activity to estimates of other sizes of rectangles and other designs was tested. The transfer effects were found to be small, unreliable, and not comparable to improvement in the specific activity practiced. Results such as these were presumptive evidence for the importance of identical elements.

Complete identity of stimuli and of responses is, of course, impossible. Abandoning the requirement of pure identity, it is usually possible to describe dimensions of similarity along which stimulus and response elements in different situations can be shown to vary. The definitions of such dimensions and the operations for their measurement must necessarily vary in different areas of investigation. The utility of a dimension of similarity depends not only on the exactitude of the measurements which it allows but, above all, on the fruitfulness of the hypotheses which can be formulated and tested with its aid. The dimensional analysis of verbal materials presented in Chapter 15 may serve as an illustration.

How elementary is an element? Behavior is a continuous process, and elements such as muscular movements and specific verbal units represent abstractions on the part of the investigator. It may be stated in defense of the theory of identical elements that its authors intended the term *element* to be used with a high degree of flexibility, according to the requirements of specific experimental prob-

lems. In some cases, such as the analysis of highly circumscribed acts, it may be useful to define *element* in terms of specific movements; in other cases, such as problem solving, an element may be conceived as broadly as "general mode of attack" or "attitude." The term element was probably unfortunate in implying reduction to minimal units, and we shall do well to follow Woodworth in speaking of *components* instead.[2]

Choice of Units in Terms of Experimental Usefulness. As long as we wish to study transfer of training (or any other kind of behavior, for that matter) experimentally, we must apply analysis into units which can be handled experimentally. We may choose such units or "components" at different levels of specificity and complexity, according to the needs and emphases of a given investigation. We may compare two tasks with respect to circumscribed stimulus and response units (granting that they may not be absolutely identical) *and/or* with respect to general attitudes and modes of attack. In the course of experimentation, we may succeed in breaking down a rough and broad unit into more specific and circumscribed ones if that is useful for prediction and control. The question, what is transferred in a transfer experiment, cannot be answered by a cut-and-dried formula. In any given experiment, we may attempt an answer in terms of specific stimulus-response associations and/or in terms of general attitudes and modes of attack.

The Experimental Analysis of Transfer

The task of the experimenter has been, of course, to define and study significant variables of which transfer is a function. The experimental work devoted to the study of such variables could be classified in several ways; for example, according to stimulus materials, type of experimental procedure, type of activity, and so on. We prefer to marshal the evidence in relation to some general principles of transfer which have arisen from the experimental work. What are the *achievements* of transfer, and in what ways are these achievements brought about?

Learning to Perform a Task. Whatever field we work in, whatever skill we are trying to acquire, there are certain general attitudes and general modes of attack without which no specific problem can

[2] R. S. Woodworth, *Experimental psychology,* New York: Henry Holt and Co., 1938, p. 177.

be solved with full success. It is through transfer that such general adjustments are learned and effortlessly applied to a manifold of situations.

As a first example, let us consider sensory and perceptual skills. Psychophysicists have long been in the habit of distinguishing between "trained" and "untrained" observers. In experiments concerned with the measurement of sensitivity, it is usually found that experienced observers have lower thresholds and give more reliable judgments than do inexperienced ones. It is extremely unlikely that the sensory systems of experienced observers have been made more acute by continued practice. A more reasonable interpretation is that the trained observer has learned how to utilize cues with maximum efficiency, to discount irrelevant indications, and to resist distracting influences. The beneficial effects of past experience on sensory capacity are thus most probably due to the transfer of observational techniques and attitudes. The extent to which perceptual performance can benefit from training (transfer) was dramatically illustrated in an experiment in which subjects were trained to recognize stimulus items presented tachistoscopically at very rapid speeds. The speed of reading of some subjects was increased on the average from 547 to 1295 words per minute. By observing and discriminating we learn how to observe and discriminate.[3]

We also become more skillful at memorizing by memorizing. For example, the more lists of nonsense syllables we learn, the more quickly we can learn new lists. The initial improvement after the first few lists is very considerable. After about ten lists, however, further improvements in memorizing skill are likely to be rather small.

The learner soon adopts specific skills and attitudes which are indispensable for the successful mastery of nonsense material. He learns to utilize rhythm and other methods of grouping, he finds out what types of associations are useful and what kinds are likely to hinder his progress. Soon the task of learning nonsense syllables no longer appears strange, and the subject approaches each new list with confidence.

The ability to perform perceptual-motor tasks derives similar

[3] S. Renshaw, The visual perception and reproduction of forms by tachistoscopic methods, *J. Psychol.*, 1945, *20*:217-232.

benefits from continued practice. When we have to trace a stylus maze for the first time, we are likely to be somewhat awkward in our movements and our "trials and errors." Both animal and human subjects show increasing speed and accuracy in the mastery of mazes with continued practice. Other perceptual-motor tasks, such as cancellation of letters, show like improvements.

The Specificity of Transfer. The attitudes and adjustments which we acquire and transfer from one situation to another may have a more or less wide range of application. Transfer effects, in short, may be *more or less specific.*

In learning how to observe, we may often learn only how to observe a circumscribed class of events. In experiments on tactual sensitivity, for example, it has been found that the two-point limen—the ability to discriminate the presence of two stimuli from one—is lowered considerably in a practiced area but only little in other areas. We have already cited experiments in which practice in estimating areas showed only little and irregular transfer to the task of estimating the areas of other designs. When the stimulus materials were objects of different weights and lines of different lengths, similar results were obtained. The fact, however, that in these experiments transfer, though small, is usually positive indicates that certain attitudes and skills *are* transferred but that they do not suffice to result in improvement as great as is achieved by direct practice on a specific problem.

The picture is similar in the fields of memorizing and perceptual-motor tasks. Practice with nonsense syllables does greatly improve our skill in memorizing syllables, but the transfer to memorizing other material, such as prose passages and poetry, tends to be small and unreliable. In some cases, indeed, the transfer effects are negative. And so it is with cancellation tests. When subjects practice cancellation of specific letters, they are hindered rather than helped by their experience when switched to canceling specific parts of speech. The specificity of transfer raises with urgency the question of specific determinants of transfer, of the factors on which sign and amount of transfer depend. With some of the determinants, we shall be concerned in the sections which follow.

Direction of Associative Change. Let a subject be trained in Task T_1 and the transfer effects of this training be tested on Task T_2.

T_1 and T_2 may differ from each other in various ways. Analyzing the two tasks into stimulus and response components, we may find that the stimuli in the two situations are (nearly) the same but that different responses are required of the subject. On the other hand, the stimulus components in the two situations may be different and the same response demanded in both situations in spite of the stimulus difference. Finally, T_1 and T_2 may, of course, differ in respect to both stimulus and response components.

Amount and sign of transfer may vary with the direction of the associative change which the switch from T_1 to T_2 entails for the subject. Must he learn to make an old response to a new stimulus or must he learn to make a new response to an old stimulus? There is good experimental evidence for the generalization that transfer effects are positive when an old response must be transferred to a new stimulus. Positive transfer effects are even greater if the stimulus is only partly changed and the old response is retained. The degree to which T_1 has been learned is, of course, an important parameter: the better practiced T_1 is, the greater are the positive transfer effects.

On the other hand, if the subject must learn to make a new response to an old stimulus, the transfer effects are likely to be negative. Indeed, the old response, which had been correct for T_1, may frequently intrude overtly during the performance of T_2. Again, the degree to which T_1 has been learned is an important parameter. The more thoroughly T_1 has been practiced, the less serious the negative transfer effects are likely to be (the sign may indeed become positive when T_1 is very well mastered).

These principles can best be demonstrated in situations which can be readily analyzed into discrete stimulus and response components. In the area of perceptual-motor tasks, the maze, with its discrete entrances and alleys, meets this need. In one transfer experiment, for example, rats were trained to withdraw from an alley in response to the appearance of a signal (light, sound, or shock). After the response had been well established, the nature of the signal was changed. For example, if the animals had been originally trained with a light, a sound or shock was substituted, and so on, for the other possible rotations. Invariably the new association was learned faster than the original one. The transfer effects are positive when an old response is attached to a new stimulus.

Verbal materials learned by the method of paired associates can also be conveniently analyzed into stimulus and response components. T_1 consists of a list of paired associates *A-B* (where *A* and *B* may be nonsense syllables or meaningful words). T_2 again is a list of paired associates: *A-C* if the stimulus is the same and the response is changed; *C-B* if the stimulus is new and the response is unchanged. In an experiment using both these arrangements, the generalization about transfer as a function of change in stimulus and in response was verified. Change in the stimulus without change in response (*C-B*) yields positive transfer; change in the response without change in the stimulus (*A-C*) tends to yield negative transfer.

The asymmetry of transfer effects produced by changes in the stimulus and changes in the response suggests that negative transfer effects may be due, at least in part, to competition among incompatible responses or "reproductive inhibition." If in T_1 the association *A-B* is established and superseded by the association *A-C* in T_2, the responses *B* and *C* are both associated with *A* and in conflict with each other. Such a conflict between incompatible tendencies inhibits, at least temporarily, the response. After prolonged practice, of course, the association *A-C* becomes stronger than the association *A-B*, and the negative transfer effects are overcome.

Equivalence of Stimuli. The physical identity or similarity of stimulus components in two or more situations is not in itself a guarantee of positive transfer effects. The identical or similar stimuli must *function* as *equivalent* for the organism in order that transfer occur. There are, indeed, many identities in nature which we can discover only after much training and painstaking analysis. Given certain conditions in the environment, equivalence depends on the organism, its capacities, needs, and attitudes. In the last analysis, it is the selective dispositions of the organism which render stimuli equivalent.

Equivalence of stimuli can be defined experimentally. If in the presence of two stimulus constellations, S_1 and S_2, behavior remains more or less unaltered, S_1 and S_2 may be considered equivalent stimuli. Starting with a constellation, S_1, we can progressively alter its characteristics along various dimensions (such as size, brightness, weight, etc.) until response to the stimulus undergoes a significant

change. The range of differences which can be introduced into S_1 without leading to a specified behavior change defines a *range of equivalence.*

It is not always possible to state in exact physical terms on what characteristics the equivalence of stimuli depends. Consider, for example, Fig. 104. Monkeys were reinforced for responding to the square (S_1) and failed to be rewarded for responding to the circle

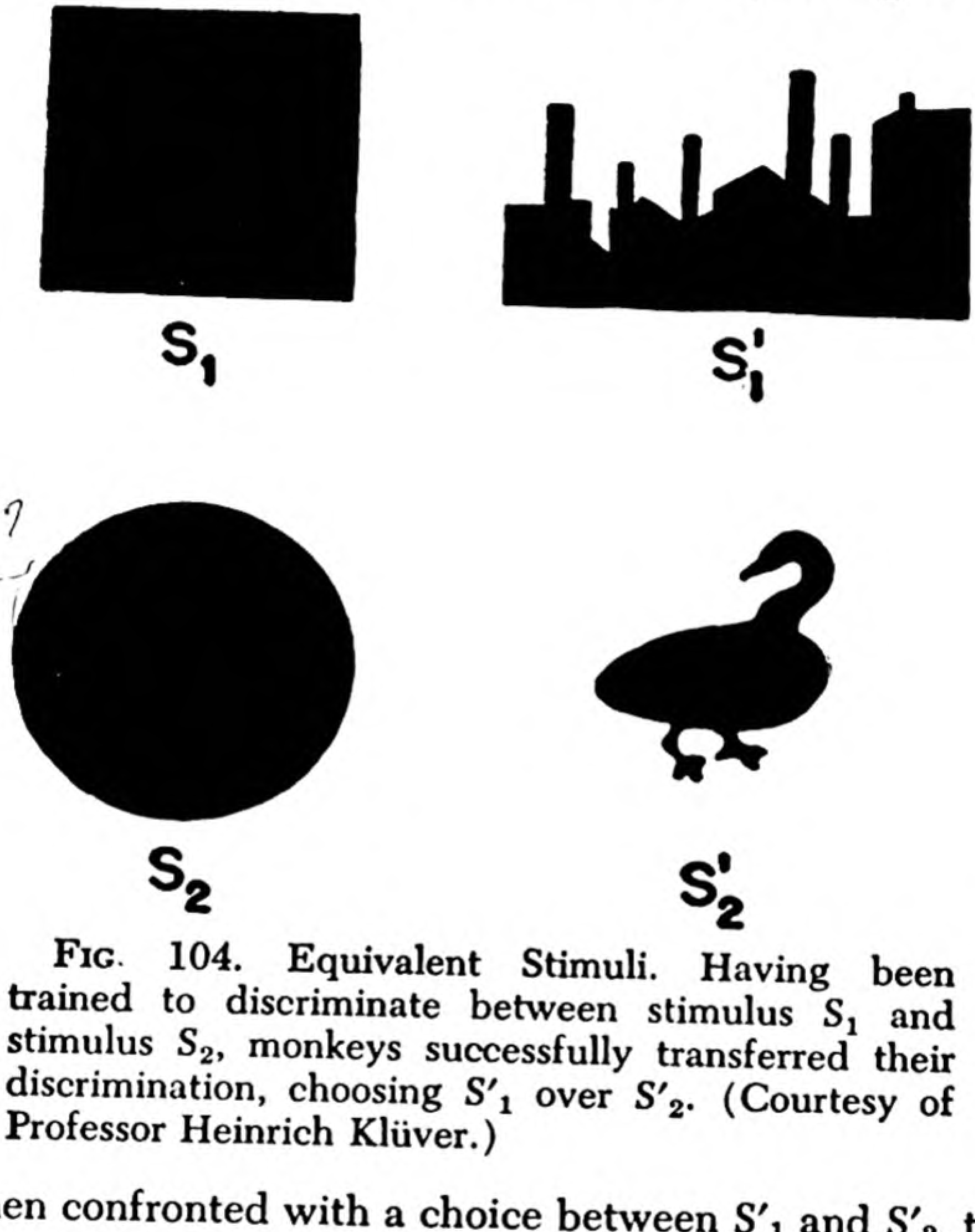


FIG. 104. Equivalent Stimuli. Having been trained to discriminate between stimulus S_1 and stimulus S_2, monkeys successfully transferred their discrimination, choosing S'_1 over S'_2. (Courtesy of Professor Heinrich Klüver.)

(S_2). When confronted with a choice between S'_1 and S'_2, they successfully transferred their discrimination, choosing S'_1 over S'_2. The difference between S'_1 and S'_2 is equivalent to the difference between S_1 and S_2, the common factor presumably being a difference along the dimension of angularity-circularity.

What renders stimulus constellations equivalent may thus be the *perception of a relationship* common to two situations. Experiments

specifically designed to show that transfer may depend on the perception of relationships common to different situations are known as *transposition experiments.* In these experiments, the physical identity of the stimuli is changed while certain relations among the stimuli are maintained. In one experiment, for example, chicks were trained to peck food from the darker of two gray pieces of cardboard. After the discrimination was thoroughly established, the chicks were given test trials during which the lighter gray was removed and replaced by an even darker shade than that of the positive stimulus. The design may be summarized as follows:

Training Trials: G_1 and G_2, where G_2 was darker than G_1
Test Trials: G_2 and G_3, where G_3 was darker than G_2

Confronted with the new stimulus constellation, the animals chose G_3 which was now the darker one, and not G_2 which was the identical shade that had been reinforced throughout the training period. Similar findings have been reported with other organisms and with other types of relationships involving relative weight, size, and sound characteristics. The transposition experiments serve to emphasize the fact that similarity or identity cannot be defined always in terms of *sameness* of physical stimulation, but rather may refer to equivalence of relationships which the organism recognizes in the environment.

Transfer as Failure to Discriminate. Transfer then may depend on the perception of *common* features in different situations. Transfer effects may, however, also be due to a failure to perceive different situations as different. The phenomenon of sensory generalization in conditioning, which is a type of transfer effect, will serve to illustrate this point.

When a conditioned response is established, the response is frequently elicited, not only by the specific stimulus used in the training but also by stimuli which differ more or less from the training stimulus. In one conditioning experiment, for example, the galvanic skin response was conditioned to a tone of fixed pitch. Tones which differed in pitch from the sound used in the training also evoked the galvanic skin response. The amplitude of the response, however, decreased with the difference in pitch between training stimulus and test stimulus. A similar *gradient of generalization* was obtained

when training tones and test tones differed in loudness. The generalization may be considered as a failure to discriminate between the stimulus used in the training procedure and the various test stimuli which elicited the response.

The phenomenon of sensory generalization is not limited to the conditioning situation. It may also be demonstrated in learning by the method of paired associates. When, for example, geometrical designs are associated with letters, alterations in the designs decrease but do not eliminate the letter responses. Again, the greater the alterations in the stimulus designs, the less frequently do they evoke the associated responses. When the terms associated are nonsense syllables and/or meaningful words, analogous results are obtained.

Experiments such as these highlight the fact that transfer from one situation to another may sometimes stem from a failure of discrimination. Transfer effects resulting from a failure to discriminate may sometimes be corrected by further training. In the case of sensory generalization in conditioning, training may reduce the amount of transfer. By differentially reinforcing only responses to a specific stimulus and failing to reinforce responses to similar stimuli, the range of generalization may be reduced effectively. The net amount of transfer will be determined, then, by the summation (or perhaps more complex interaction) of generalizing and discriminating (inhibitory) tendencies.

Practice With and Without Insight into Principles. The activity T_1 *from which* there are transfer effects to an activity T_2 may be practiced with and without explicit understanding of the principles on which successful transfer would depend. A classical experiment illustrates this point.

A group of boys were given the task of hitting a target 12 inches underwater. To master the task, they had to learn to take into account the displacement of the target due to refraction of light. The correction for refraction varies, of course, with the distance from the surface. One group of subjects was explicitly taught the principle of refraction; the other group remained uninstructed. Both groups improved equally on the 12-inch target. When the distance of the target from the surface was changed to 4 inches, the instructed group showed considerable positive transfer effects, the uninstructed group hardly benefited at all from its previous practice.

The superior transfer effects achieved by the study of principles as compared with rote learning has also been demonstrated in the field of memory. Two groups spent identical amounts of time in memory practice. One group spent the entire period in memorizing, whereas the other group divided its time between memorizing and the study of effective *methods* of memorizing. In the rote group there were hardly any transfer effects to new memory tasks, the methods group benefited greatly from the training period and showed substantial positive effects. The same conclusion has been reached with a variety of other learning tasks: mazes, ideational problems, school subjects, to name but a few. It is true that we learn by doing, but we learn best by performing tasks "insightfully," i.e., with full awareness of the principles guiding successful perfoimance.

Cross-Education

Some of the general principles which we have discussed are exemplified in a special type of transfer: cross-education.

The Nature of Cross-Education. Practicing an activity with a particular part of the body usually facilitates performance of the same activity with another part of the body. Such, in brief, is the definition of cross-education. Usually the positive transfer effects are between symmetrically located parts of the body, from eye to eye, from hand to hand, and from foot to foot. When cross-education is from one body organ to its symmetrical counterpart, we speak of *bilateral transfer*. Although it is by far the most common type of cross-education, bilateral transfer is not the only kind. Thus there may be transfer from hand to foot, or even from one skin area to another, nonsymmetrical one.

Reference Experiments. Cross-education, and bilateral transfer in particular, has been demonstrated in a great variety of situations, especially with perceptual-motor activities. In a typical experiment on cross-education, the task remains constant throughout, but the body parts with which it is performed are varied. We shall illustrate with one of the most common experiments in the study of cross-education: bilateral transfer in mirror tracing.

The subject's task is to trace a pattern, such as the star shown in Fig. 105, which is hidden from direct view, guiding his movements by the reflection of the pattern in a mirror. A variation on

this procedure requires the subject to aim at, and hit, a target reflected in a mirror. This task is initially quite difficult for many subjects. It may take quite a time before the reversals in spatial direction introduced by the mirror are fully taken into account. The task lends itself ideally to the study of bilateral transfer because (1) it is a task which can be well performed with symmetrical

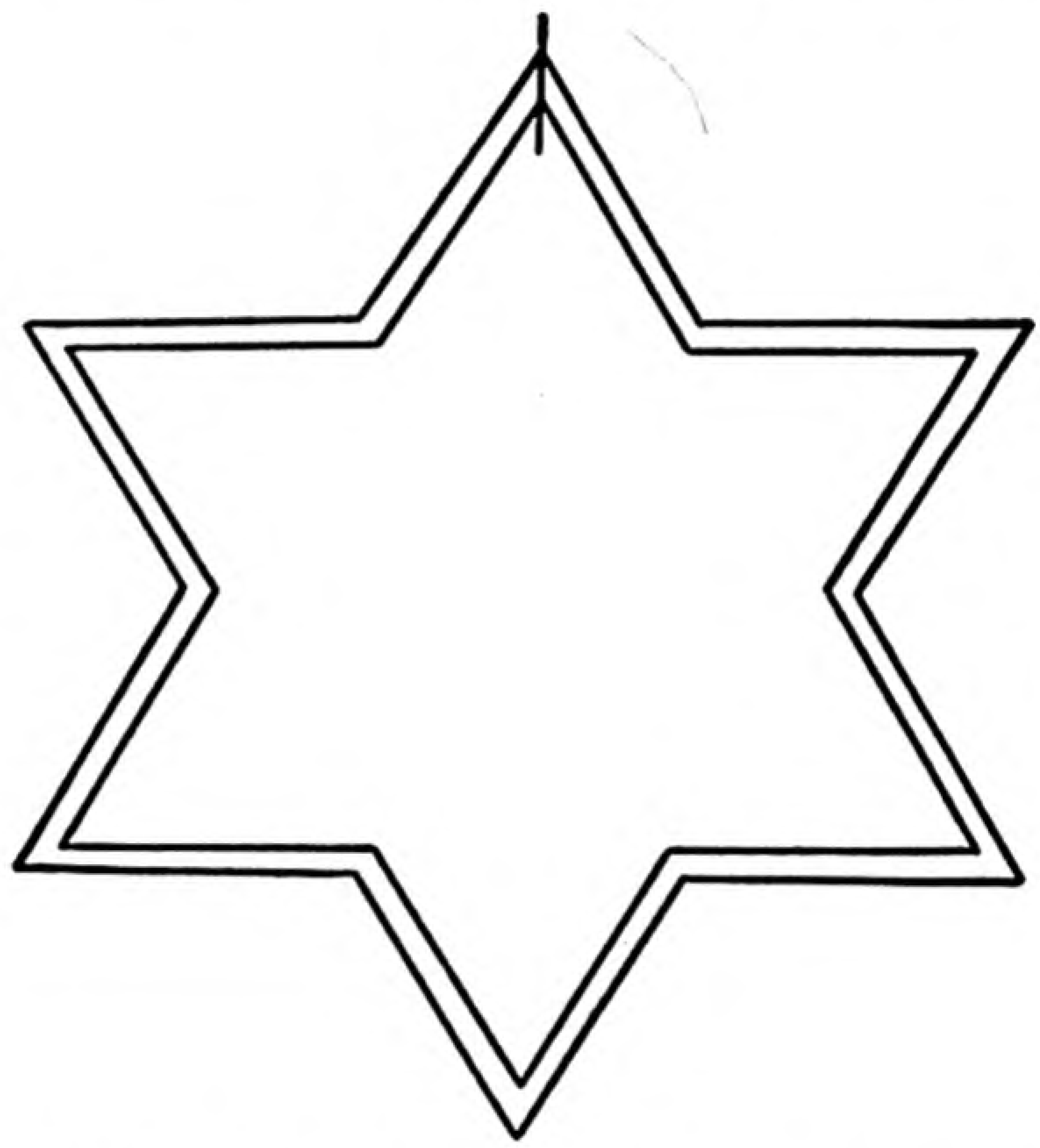


FIG. 105. Star Pattern Used in Experiments on Bilateral Transfer of Training in Mirror Tracing.

members—hands or feet—and (2) most subjects start at a low level of skill but improve considerably after some practice.

One conventional design uses an experimental group and a control group. The experimental group receives a foretest with the nonpreferred member (e.g., the left hand), a training period with the preferred member (e.g., the right hand), and an aftertest with the nonpreferred member. The control group receives only the foretest and aftertest. To the extent that the experimental group shows a greater gain from foretest to aftertest, there has been positive bilateral transfer. Use of a control group, of course, is necessary to take into account the practice provided by the foretest.

The results of cross-education experiments have uniformly shown positive transfer effects, even from hand to foot. Fig. 106 graphically shows the phenomenon of transfer from hand to foot in mirror aiming. The experimental group was given sixty trials with the right hand while the control group rested. The graph shows the course

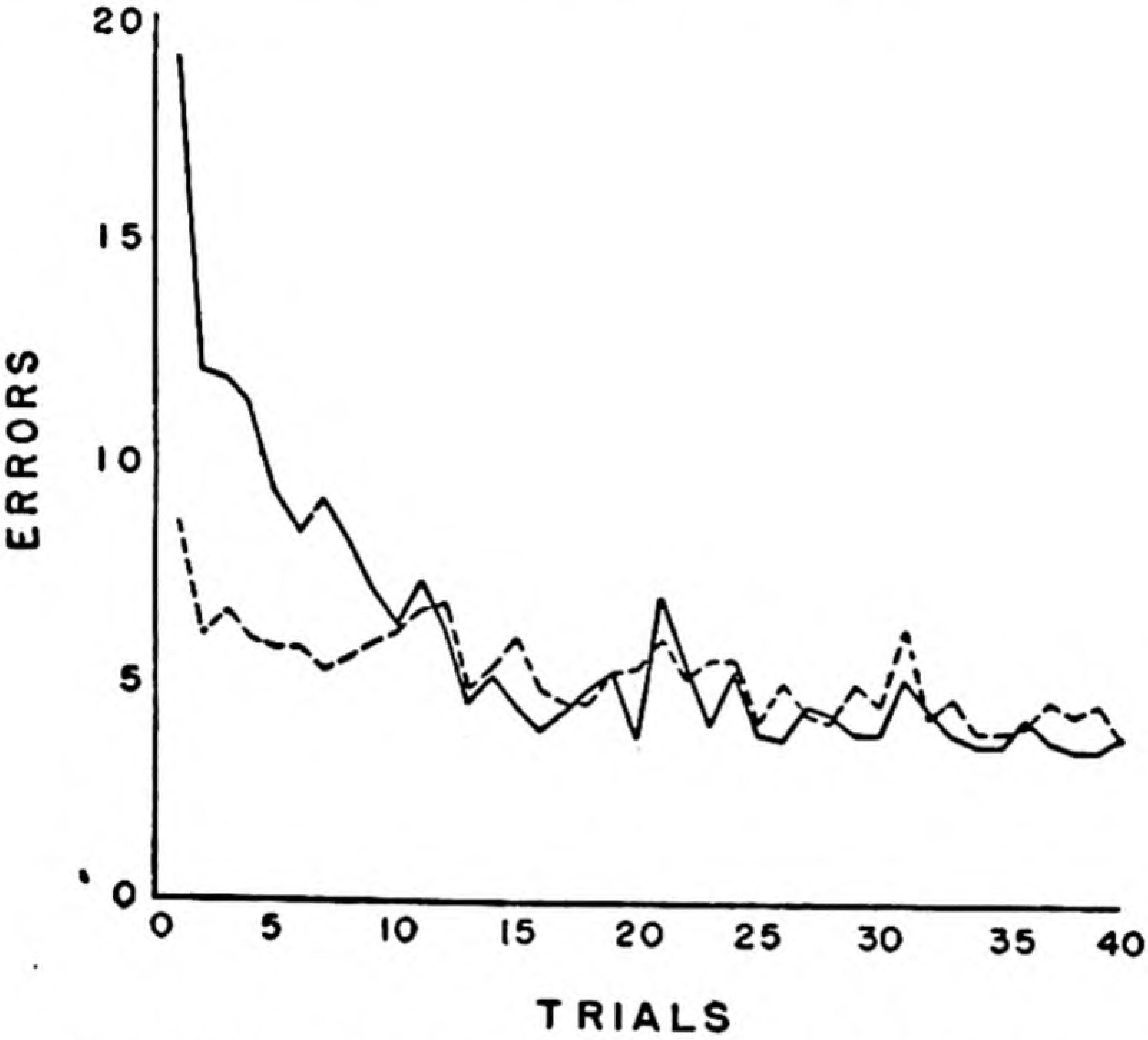


FIG. 106. Transfer in Mirror Aiming. The dotted line represents the performance of the experimental group; the solid line, the performance of the control group. (From C. W. Bray, Transfer of learning, *J. Exper. Psychol.*, 1928, *11*:450, by permission of the journal and the American Psychological Association.)

of performance with the right foot by the two groups. The initial advantage of the experimental group is striking: members of this group make only half as many errors as the controls. However, the transfer group maintains its advantage only over the first block of ten trials. Thereafter, the two curves overlap so much as to be almost indistinguishable. It is rather typical of this kind of experiment that the advantages (or disadvantages) resulting from transfer are confined to the early parts of the aftertest. If it is primarily

modes of attack and attitudes which are transferred, we may infer that the control group makes appropriate adjustments and catches up with the transfer group at a rather early stage.

Cross-education has been demonstrated with many activities besides mirror tracing and mirror aiming. To name but a few other perceptual-motor tasks, we have: tracing a maze pattern, ball tossing, tapping, pursuit of rotating target, and so on. The phenomenon is well established, and to multiply the demonstrations much more would have little value.

Cross-education is, of course, not limited to motor activities. In the sensory field, we have come to take cross-education so much for granted that its experimental demonstration seems almost unnecessary. Close one eye and look at a figure. Then close the other eye and look at the same figure. Of course you recognize it. The result seems so obvious that the demonstration appears most trivial. Yet, it is an example of bilateral transfer whch occurs constantly in the development of perceptual skills. If the perceptual task is made more difficult, the fact of bilateral transfer has seemed less obvious and has been considered deserving of experimental demonstration.

In one experiment, for example, visual recognition was complicated by exposing the objects peripherally. The subjects needed a certain amount of training in order to acquire proficiency in this perceptual task. After practice with one eye, there was 100 percent transfer to the other eye! Similar results have been obtained in experiments on tactual sensitivity. Improvement in discriminating one-point stimulation from two-point stimulation (the two-point limen) yields virtually complete transfer effects to corresponding areas of the skin, e.g., from one hand or arm to the other. Having learned how to read Braille with one hand, subjects can read it almost as well with the other.

As we have already pointed out, cross-education is not confined to bilateral transfer. In mirror aiming, for example, there is transfer from hand to foot, both homolaterally and heterolaterally. There is abundant evidence for *intersensory* transfer. What we learn with the aid of the sense of hearing, we can apply in tasks which are primarily visual, and vice versa. We continually apply what we have learned about objects by vision to tactual manipulation. Indeed,

the coöperation of visual and tactual responses has long been believed central to the development of space perception. Such a theory necessarily implies transfer across modalities. There is, indeed, abundant evidence that cross-education greatly increases the repertoire of the organism's responses and the flexibility of its adjustment to the environment.

The Lesson of Cross-Education. The fact of cross-education serves to drive home a conclusion to which we have already pointed. Transfer from one situation to another does not depend on the physical identity of stimuli or on the identity of responses. When there is transfer from one eye to the other, different pathways of stimulation are necessarily involved, and so for other types of sensory transfer. When a skill is transferred from one limb to another, different muscle groups are necessarily used. Yet, cross-education in general, and bilateral transfer in particular, are among the most common and most reliable types of transfer.

Clearly, stimuli function as equivalent for symmetrical sense organs, and different parts of the body are more or less equivalent on the response side. Frequently the equivalence lies in the fact that general principles, such as correction for mirror distortion or methods of attacking a maze problem, can be put into effect by more than one combination of response patterns. About the neurological conditions of such transfer, we can say little, except to point to the obvious implication that the effects of practice cannot be limited to a few specific pathways and projection areas exercised during the training.

The important task for the experimenter is to study ranges of equivalence and their determinants. What kinds of stimulus situations evoke similar response patterns? What kinds of response patterns are equivalent in the results which they achieve for the organism?

Experiment XXV

Transfer of Training in Maze Learning

Purpose. To demonstrate transfer of training in maze learning; specifically, to show that learning of a maze facilitates the learning of a new maze.

Materials. Two stylus mazes, stylus, blindfolding goggles or cloth,

stop watch, and record sheets are required. A *stylus maze* is a maze whose alleys consist of grooves which the subject traces by means of a stylus.[4] If possible, the two mazes should be constructed so that some parts are identical and some parts are different: some alleys are correct for both mazes, and some alleys are correct for one but blind for the other. However, the experiment can be performed adequately without such structural overlap of the mazes.

For most efficient recording, the score sheets should have a plan of the maze so that the experimenter can easily enter the subject's moves. Some mazes are equipped with sheets of carbon and white paper underneath the grooves so that the subject's moves are automatically recorded. Since such carbon tracings are sometimes difficult to read, the experimenter should always keep a record in addition.

Design of the Experiment. For this experiment, members of the class form groups of three. One serves as the experimenter and two are subjects. In this case, it is not advisable for the same person to serve first as experimenter and then as subject, since watching the performance of another provides an uncontrolled amount of information about the maze pattern. With two experimental subjects, we use one of the designs discussed above.

Subject 1: Maze 1, followed by Maze 2
Subject 2: Maze 2, followed by Maze 1

This design renders it unnecessary to equate strictly the ability of the subjects or the difficulty of the two mazes.

Procedure. The subject is seated in front of the maze, blindfolded. He is given the stylus and his hand is put at the entrance of the maze. The experimenter instructs him to work as rapidly but also as accurately as possible. The subject begins his first trial at a signal from the experimenter. The experimenter starts his stop watch at the moment of giving the signal and carefully times all the trials thereafter.

Shortly (say, 30 seconds) after the first trial, the second trial is started. The trials are continued until the subject has reached the criterion of mastery. A criterion of either one, two, or three errorless runs may be used.

After mastering the first of the mazes (Maze 1 for Subject 1 and Maze 2 for Subject 2), the subject is given a rest period of, say, 10 minutes. He then begins to learn the second maze (Maze 2 for Subject 1 and

[4] Instead of a stylus maze, a *finger maze* may be used. The alleys here consist of wires which the subject traces with his fingers.

Maze 1 for Subject 2). The procedure is in all respects identical with that used before, and the trials are again continued until the criterion of mastery is reached. As we have indicated, the duration of each trial is timed by means of a stop watch.

How to Score the Subject's Performance. If the record sheet contains a plan of the maze, the experimenter can enter the subject's moves as he sees them being made. On that basis, the number of errors can be computed for each trial. If the record sheet does not feature a maze plan, the experimenter may tally the number of errors. The latter procedure has, of course, the disadvantage that the location of the errors cannot be identified.

An *error* is conventionally defined as (1) entrance into a blind alley, (2) retracing in the correct path. In scoring the subject's performance, the experimenter should have a clear definition of *error* in mind. He must decide, for example, how far the stylus must have been moved into a blind alley in order to be classified as an entrance.

Treatment of Results. The critical comparison is between performance on the first task and performance on the second task. According to our counterbalanced design, the performance of the two subjects on the two mazes is, thus, pooled.

We compare the performance on the task learned first and on the task learned second with respect to the following scores:

1. Number of trials required to reach criterion
2. Total amount of time required to reach criterion
3. Total number of errors made before criterion was reached

To trace the temporal course of the performance, we obtain:

4. Number of errors on each individual trial
5. Amount of time spent on each individual trial

Learning curves are obtained by plotting time and errors against trial number.

Comparison of the two maze performances with respect to number of trials, number of errors, and amount of time required to reach criterion, allows us to test for transfer effects. If there is a significant difference in favor of the second task, the transfer effects are positive; if the difference is in favor of the first task, the transfer effects are negative. Finally, if there is no significant difference, there is zero transfer. It is interesting to compare and contrast the transfer effects revealed by the three measures. Does one of them show greater effects than the others?

To quantify the *amount* of transfer, we compute the percentage of *saving* attributable to transfer (see p. 356). Suppose it took 12 minutes

to master the first task and only 8 minutes to master the second task. The difference is 4 minuces or 33.3 percent of 12. We conclude that the amount of transfer, as measured by time scores, is 33.3 percent. Similar saving scores can be computed for trials and errors.

To represent the transfer effects graphically, we plot the learning curves (Vincent curves may have to be used in combining the results of the two subjects; see pp. 284-286). If the time and error scores for the second task fall more rapidly than for the first, this difference may be ascribed to positive transfer effects.

It may be instructive to compare scores for specific alleys. If the two mazes are structurally similar, some responses could have been directly transferred (the same alley being correct in both cases), and some responses had to be changed (an alley being correct in one maze and blind in the other). It may well be that there are negative transfer effects for specific turns although the overall effects are positive.

Finally, it may be useful to examine the subject as to the methods he employed in mastering the mazes. Did he use verbal self-instructions such as "first left, then right, then left again," etc.? Did he rely on a visual scheme of the maze? To what extent did he explicitly verbalize the principles which made positive transfer effects possible? A careful examination of the subject's report may help to throw light on the factors responsible for the transfer effects.

Experiment XXVI

Bilateral Transfer in Mirror Tracing

Purpose. To measure the transfer from practicing a novel task with one hand to performance of the same task with the other hand.

Materials. A typical arrangement for mirror tracing consists of a board to which the sheet with the star pattern (Fig. 105) can be tacked, a mirror in which the pattern is reflected, and a hand shield which prevents the subject from looking directly at the pattern. A stop watch is needed to time the trials.

Experimental Design. The experimental design calls for an experimental group and a control group. The members of both groups receive a foretest in which they trace the pattern with their nonpreferred hands. The members of the experimental group then are given twenty trials with their preferred hands while the controls rest. Both groups finally receive an aftertest in which they trace the pattern with their nonpreferred hand.

Procedure. The subject is seated in front of the apparatus. The hand shield prevents him from looking at the star directly, but he can see the

non-preferred

pattern reflected in the mirror. The subject slips his hand under the shield and places the point of his pencil on the starting point. The starting point is directly opposite the subject (i.e., nearest to the mirror). At a signal from the experimenter, the subject starts tracing the star pattern, keeping as well as he can within the lines. The experimenter times the duration of each trial with his stop watch. Throughout the trial, the experimenter watches the subject's work closely and records all errors. Touching or crossing a line constitutes an error. If the subject goes outside the line, he must reënter at the same point; otherwise, the recrossing constitutes another error.

Treatment of Results. Each trial is scored in terms of (1) the number of errors, and (2) the time required for completion. To show the course of practice with the preferred hand, we plot the number of errors and time scores against trial number. In this manner, we obtain learning curves for the preferred hand.

To gauge the amount of transfer, we determine for both groups: (1) the gain (or loss) in errors, and (2) the gain (or loss) in time between foretest and aftertest. The differences between the gains (or losses) of the two groups provide us with a measure of transfer. The significance of these differences should be tested.

Certain additional observations are of interest. How do the changes in time compare with the changes in error scores? If one measure shows transfer effects to a larger extent than the other, what might be the explanation of this discrepancy? If there is such a divergence between the two measures, how is it related to the shape of the learning curves for time scores and error scores?

General References

Katona, G., *Organizing and memorizing: studies in the psychology of learning and teaching*, New York: Columbia University Press, 1940.

McGeoch, J. A., *The psychology of human learning*, New York: Longmans, Green and Co., 1942, Chap. 10.

Orata, P. T., *The theory of identical elements, being a critique of Thorndike's theory of identical elements and a re-interpretation of the problem of transfer of training*, Columbus: Ohio State Universtiy Press, 1928.

Woodworth, R. S., *Experimental psychology*, New York: Henry Holt and Co., 1938, Chap. 8.

Design of Transfer Experiments

See General References, espccially Woodworth and McGeoch

What Is Transferred in Transfer of Training

Allport, G. W., *Personality: a psychological interpretation,* New York: Henry Holt and Co., 1937, Chap. 10.

Cook, T. W., Studies in cross education: V. Theoretical, *Psychol. Rev.,* 1936, *43*:149-178.

Sandiford, P., *Educational psychology: an objective study,* New York: Longmans, Green and Co., 1928.

Thorndike, E. L., *The psychology of learning,* New York: Teachers College, 1913.

Thorndike, E. L., and Woodworth, R. S., The influence of improvement in one mental function upon the efficiency of other functions. (I.); II. The estimation of magnitudes; III. Functions involving attention, observation and discrimination, *Psychol. Rev.,* 1901, *8*:247-261; 384-395; 553-564.

Experimental Analysis of Transfer

Learning to Perform a Task and Specificity of Transfer

Kline, L. W., Some experimental evidence in regard to formal discipline, *J. Educ. Psychol.,* 1914, *5*:259-266.

Martin, M. A., The transfer effects of practice in cancellation tests, *Arch. Psychol.,* 1915, No. 32.

Reed, H. B., A repetition of Ebert and Meumann's practice experiment on memory, *J. Exper. Psychol.,* 1917, *2*:315-346.

Renshaw, S., The visual perception and reproduction of forms by tachistoscopic methods, *J. Psychol.,* 1945, *20*:217-232.

Sleight, W. G., Memory and formal training, *Brit. J. Psychol.,* 1911, *4*:386-457.

Woodrow, H., The effect of type of training upon transference, *J. Educ. Psychol.,* 1927, *18*:159-172.

Direction of Associative Change

Bruce, R. W., Conditions of transfer of training, *J. Exper. Psychol.,* 1933, *16*:343-361.

Gibson, E. J., Sensory generalization with voluntary reactions, *J. Exper. Psychol.,* 1939, *24*:237-253.

Siipola, E. M., Implicit or partial reversion-errors: a technique of measurement and its relation to other measures of transfer, *J. Exper. Psychol.,* 1940, *26*:53-73.

Wylie, H. H., An experimental study of transfer of response in the white rat, *Behav. Monogr.*, 1919, *3*, No. 16.

Yum, K. S., An experimental test of the law of assimilation, *J. Exper. Psychol.*, 1931, *14*:68-82.

Equivalence of Stimuli

Allport, G. W., *Personality: a psychological interpretation*, New York: Henry Holt and Co., 1937, Chap. 10.

Klüver, H., *Behavior mechanisms in monkeys*, Chicago: University of Chicago Press, 1933.

Köhler, W., Optische Untersuchungen am Schimpansen und am Haushuhn, *Abn. Preuss. Akad. Wiss., Phys.-math. Kl.*, 1915, No. 3.

Transfer as a Failure to Discriminate

Hovland, C. I., The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone, *J. Gen. Psychol.*, 1937, *17*:125-148.

Hovland, C. I., The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone, *J. Genet. Psychol.*, 1937, *51*:279-291.

Lashley, K. S., and Wade, M., The Pavlovian theory of generalization, *Psychol. Rev.*, 1946, *53*:72-87.

Spence, K. W., The differential response in animals to stimuli varying within a single dimension, *Psychol. Rev.*, 1937, *44*:430-444.

Practice With and Without Insight into Principles

Bagley, W. C., *The educative process*, New York: The Macmillan Company, 1905.

Cox, J. W., Some experiments on formal training in the acquisition of skill, *Brit. J. Psychol.*, 1933, *24*:67-87.

Judd, C. H., The relation of special training to general intelligence, *Educ. Rev.*, 1908, *36*:28-42.

Ruediger, W. C., The indirect improvement of mental function through ideals, *Educ. Rev.*, 1908, *36*:364-371.

Woodrow, H., The effect of type of training upon transference, *J. Educ. Psychol.*, 1927, *18*:159-172.

Cross-Education

Bray, C. W., Transfer of learning, *J. Exper. Psychol.*, 1928, *11*:443-467.

Cook, T. W., Studies in cross education: I. Mirror tracing the star-shaped maze, *J. Exper. Psychol.*, 1933, *16*:144-160.

Franz, S. I., Studies in cerebral function: VIII. Training in touch perception and cross-education, *Univ. Calif., Los Angeles, Publ. Educ., Philos., Psychol.*, 1933, *1*:121-128.

Franz, S. I., and Layman, J. D., Studies in cerebral function: I. Peripheral retinal training and practice transfer, *Univ. Calif., Los Angeles, Publ. Educ., Philos., Psychol.*, 1933, *1*:65-78.

Munn, N. L., Bilateral transfer of learning, *J. Exper. Psychol.*, 1932, *15*:343-353.

» 19 «

EMOTIONAL BEHAVIOR

THE NATURE OF EMOTIONAL STATES

One of the concepts which psychology has taken over from common usage but has often found difficult to handle analytically and experimentally is that of *emotion*. We all have, at one time or another, observed organisms in a highly aroused or highly depressed state and have described such states as emotional. Our language is replete with terms thought to describe different emotional states. Many attempts have been made over the years to correlate such states with specific situations on the one hand, and with specific bodily changes on the other. The results of these investigations have often been contradictory or, at best, inconclusive.

There is no doubt that under conditions of extreme stress radical deviations from the normal level of function occur. Thus the physiologist Cannon was able to demonstrate in a long series of experiments that in emergency situations, there occurs a widespread syndrome of specific reactions. These reactions involve chiefly the circulatory, respiratory, and digestive systems and are mediated primarily by the autonomic nervous system and a hormone of the adrenal glands. Reliable as these reactions are in extreme situations, abundant experimentation has shown that there is little or no correlation between specific stress situations and specific bodily syndromes. Thus in anger, fear, general excitement, etc., highly similar bodily reactions take place which serve to mobilize the energy resources of the organism to cope with the stressful situation. These reactions are manifold, but some of the most striking changes include heightened blood pressure, greater readiness of the blood to clot when the body is wounded, slowing down of digestive activities, increase in the blood sugar level, a faster pulse, and general

increase in the tension of the skeletal muscles. But none of these changes seems to be specific to any particular stress situation.

Differentiation of Emotional States

How is it possible, then, to differentiate in language and action, with some degree of success, among the different emotional states? The answer seems to lie in the fact that we do not make our differentiations solely on the basis of bodily expressions. Our judgment of emotion is aided by at least partial knowledge of the stimulus situation and the knowledge of the type of reaction our culture expects. Thus when we witness a severe insult, we ordinarily expect the emotion of anger to be displayed. We are further aided in our judgment by our inference as to what the purpose of the overt behavior in a given situation might be, whether it be directed toward flight or attack, for example. Whatever information we may have about an individual's habits of response will also help in the interpretation of emotional behavior.

The development of a set of terms designating different emotions is also anchored in our own responses to stressful situations. Faced with various stressful situations, we have learned to employ different verbal responses. When we are insulted, we are prone to utter invectives and to describe ourselves as angry. This continuous use of labels to describe our own reactions contributes to the formation of a highly differentiated language of emotions.

It is clear, then, that what is ordinarily called *emotion* stands for a number of diverse elements: the nature of the stimulus situation, our verbal responses to it, a series of internal bodily changes, overt action in the situation, and culturally determined expectations.

Some Measures of Physiological Change in Emotion

In spite of the lack of correlations with specific situations, bodily changes form an important part of our reactions to emotional stimuli. Therefore, these bodily changes have, in fact, been widely investigated. By the nature of the problem, most of these investigations are properly conducted in the physiological laboratory. Certain of these bodily expressions can, however, be profitably included in the investigation of the behavioral and judgmental components

of emotional reactions. Thus respiratory changes, changes in blood pressure and heart rate, and changes in the resistance of the skin have been frequently measured in psychological studies of emotion. In the discussion which follows we will emphasize the experimental procedures by which such measurements are made.

Respiratory Changes. Two important phases of the respiratory cycle are distinguished: the inhalation and the exhalation of air. The temporal sequence of inspiration and expiration is described in Fig. 107. Inspiration is typically characterized by a rapid intake of air and is followed by a more gradual expiration. Both the relative amplitude and the relative duration of these two aspects of the

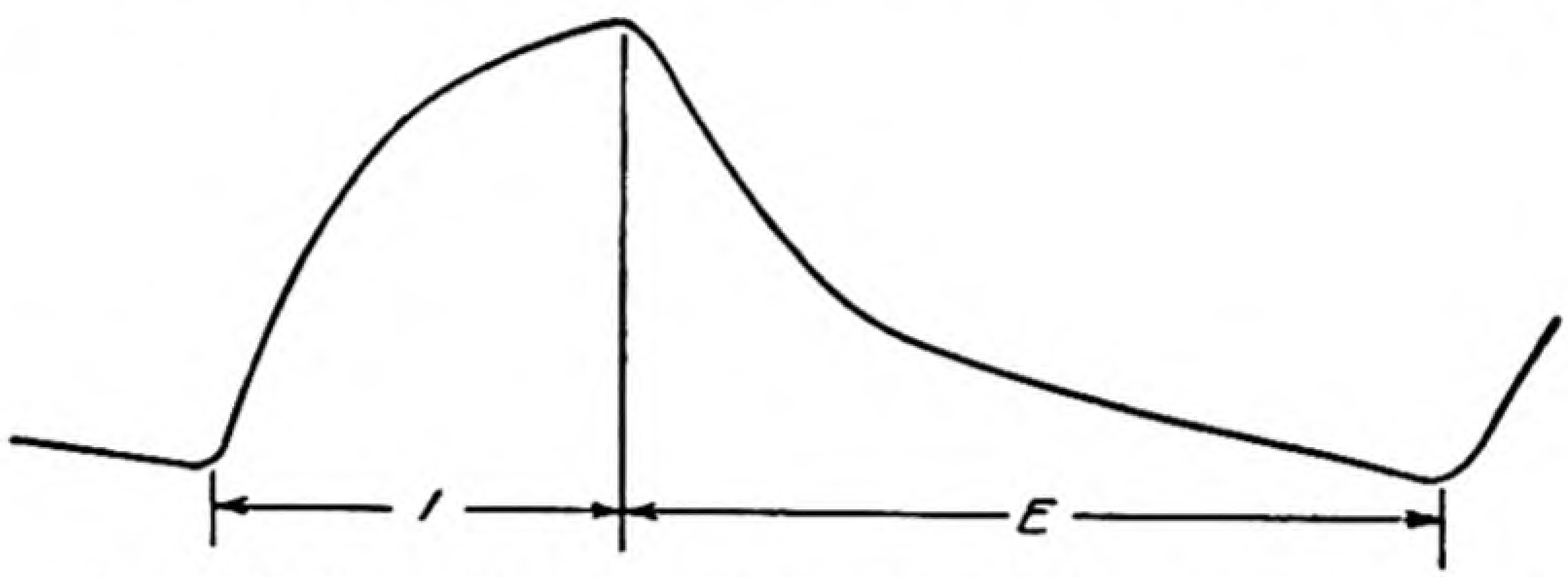


FIG. 107. The Respiratory Cycle. *I* denotes inspiration; *E* stands for expiration. (From R. S. Woodworth, *Experimental psychology*, 1938, p. 262, by permission of Henry Holt and Company, Inc.)

breathing cycle have been used in quantitative investigations. The amplitude refers, of course, to the height of the curve in Fig. 107. The temporal aspect has been quantified by the ratio of duration of inspiration to duration of expiration. This ratio, *I/E,* is called the inspiration-expiration ratio.

A device to record the durations of inspiration and expiration is called a pneumograph. This instrument consists simply of a rubber tube fastened around the chest of the experimental subject. This tube, through an airtight joint, continues to a small chamber of air. Stretched across this chamber is a rubber diaphragm. Movements of this diaphragm cause a lever to record on a revolving drum the rise and fall of the subject's chest. The combination of rubber diaphragm, attached lever, and chamber of air is known as a *tambour.* The lever is caused to move by chest movements because the system is an airtight one. The sensitivity of the *I/E* ratio

can be demonstrated in many a variety of situations. Many sudden or tension-producing stimuli cause us to "catch our breath" and thus alter the *I/E* ratio. An unexpected loud sound, a threat, the expectation of a noxious stimulus—all these produce measurable changes in the breathing cycle. Similarly, when we pay close attention to a stimulus or perform a difficult discrimination, we alter our breathing. Such activities as speaking and laughing clearly introduce radical changes. Because the *I/E* ratio is an indicator of expectation and tension, attempts have been made to use it for the discovery of deception. One experimenter found that an individual's *I/E* ratio behaved differently after he told a falsehood than after he told the truth. These results have not been generally confirmed, however, and the use of breathing changes for lie detection is still in doubt.

Changes in Blood Pressure. Variations in blood pressure have often been reported as characteristic of emotional reactions. Measurements of blood pressure are made by means of a sphygmomanometer. This instrument, shown in Fig. 108, consists of an inflatable rubber bag which is fitted around the subject's upper arm. This bag is connected to one side of a U-shaped tube which contains mercury. The upper portion of the other side of the U is evacuated. When the bag is inflated, it exerts a pressure in the U tube which causes the mercury to rise in the evacuated side. This pressure is increased until it exceeds that of the blood. In this way, the blood stops flowing through the artery. By means of a valve, this pressure is slowly reduced until the experimenter can just detect the reappearance of the pulse with the aid of a stethoscope. The pressure at this point is then indicated by the height of the mercury column. The pressure so read is known as the *systolic blood pressure.* By allowing more air to escape from the rubber bag, the pressure is further reduced until the sounds heard in the stethoscope suddenly drop in intensity and acquire a muffled quality. The pressure at this point is again read and is then known as the *diastolic pressure.*

Usually, changes in the systolic pressure are measured in emotional situations. The most general statement which can be made is that systolic pressure tends to rise during excitement. Our previous statement that specific bodily changes show little or no correlation with specific emotional situations certainly holds true in the case

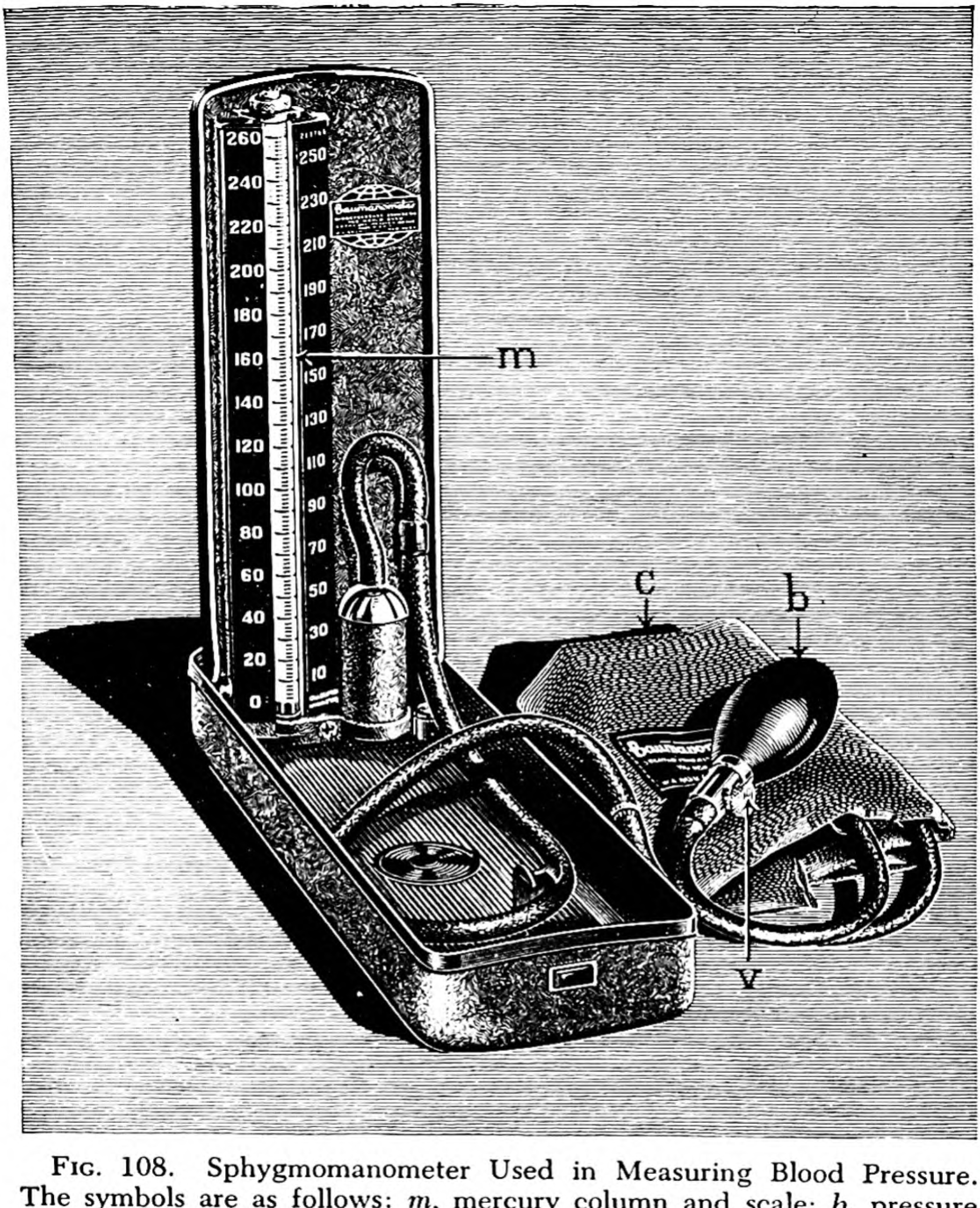


FIG. 108. Sphygmomanometer Used in Measuring Blood Pressure. The symbols are as follows: *m*, mercury column and scale; *b*, pressure bulb; *v*, needle valve; *c*, cuff. (From J. P. Fulton, *Howell's textbook of physiology*, 1946, p. 672, by permission of W. B. Saunders Co. and the W. A. Baum Co.)

of blood pressure. Changes have been recorded in a large variety of experiments. It is not possible to state what specific emotional characteristics of a situation determine a rise in blood pressure.

Blood pressure does not provide the only measure of circulatory reactions in emotional situations. Changes in blood distribution and in pulse rate, for example, are other variables of circulatory response. Special instruments are available for the measurement of each of these types of response, e.g., the *plethysmograph* for gauging changes in the volume of the blood in part of the body and the *sphygmograph* for the determination of pulse rate. As in the case of blood pressure, the general trend of the results indicates changes during excitement, but practically no correlation with specific emotional situations.

Changes in Skin Resistance. Among the most widely investigated bodily expressions in emotional situations are changes in the electrical resistance of the skin. By passing a weak current through the skin, it is possible to measure the level of skin resistance and the changes in this resistance.

For the purpose of measuring skin resistance, two electrodes (usually consisting of flat metal disks) are strapped to the subject's hand or to some other convenient part of the body. Although different types of instrumentation have been used, two items are essential: a source of current and a device for measuring resistance. Fig. 109 gives a schematic diagram illustrating the basic method of measurement. In Fig. 109 the source of current is the battery, *B*. The circle indicates the subject's hand, and the *e*'s the electrodes attached to it. The resistors in the circuit are denoted by *R*. The variable resistor, R_1, is adjusted until the galvanometer, *G*, shows no deflection. Then another variable resistor is substituted for the resistance offered by the subject's hand, and this new resistance is adjusted until the galvanometer again shows no deflection. The value of this inserted resistance matches the level of resistance of the subject's skin. For more rapid changes, it is not practical to remove the subject's hand and to substitute another resistor. To gauge rapid changes in resistance, the deflections of the galvanometer are calibrated to read skin resistance directly.

As our description of the apparatus has indicated, two types of measurement are made: measures of the general level of resistance

and measures of rapid changes. As Fig. 110 shows, at the beginning of an experimental session, the skin resistance has a relatively high value. As the experimental session progresses and the subject

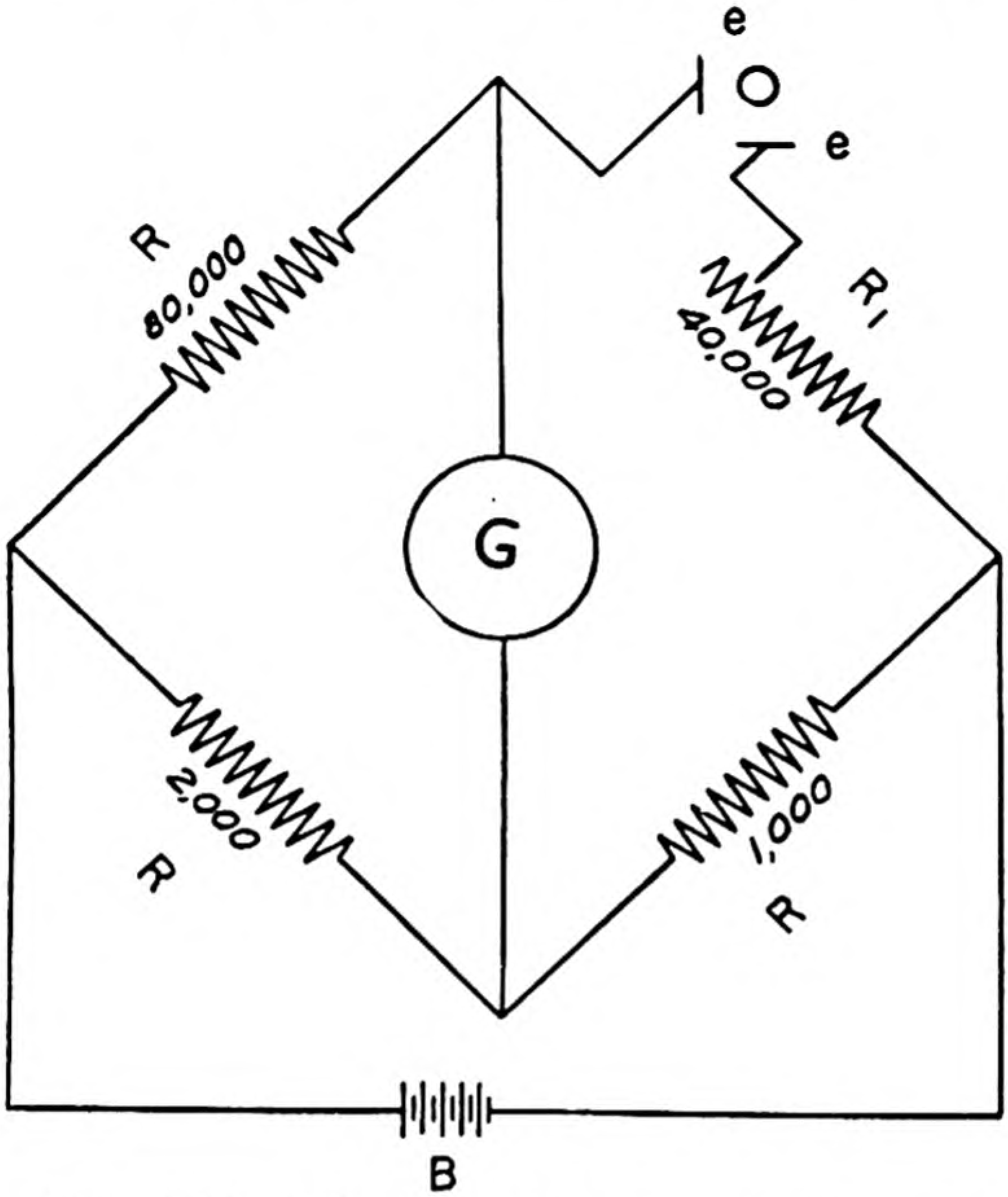


FIG. 109. Schematic Diagram of Apparatus Used in the Measurement of the Galvanic Skin Response. *B*, battery; *R*, resistor; R_1, variable resistor; *G*, galvanometer; *e*, electrode; *O*, subject. (From R. S. Woodworth, *Experimental psychology*, 1938, p. 278, by permission of Henry Holt and Company, Inc.)

becomes more apprehensive about the purpose of the electrodes, his resistance level becomes less and less (Day 1). In the case of the experiment in which Fig. 110 was obtained, a "ready" signal was given near the end of the 6th minute and was then followed by a loud noise. The heightened expectancy induced by the "ready"

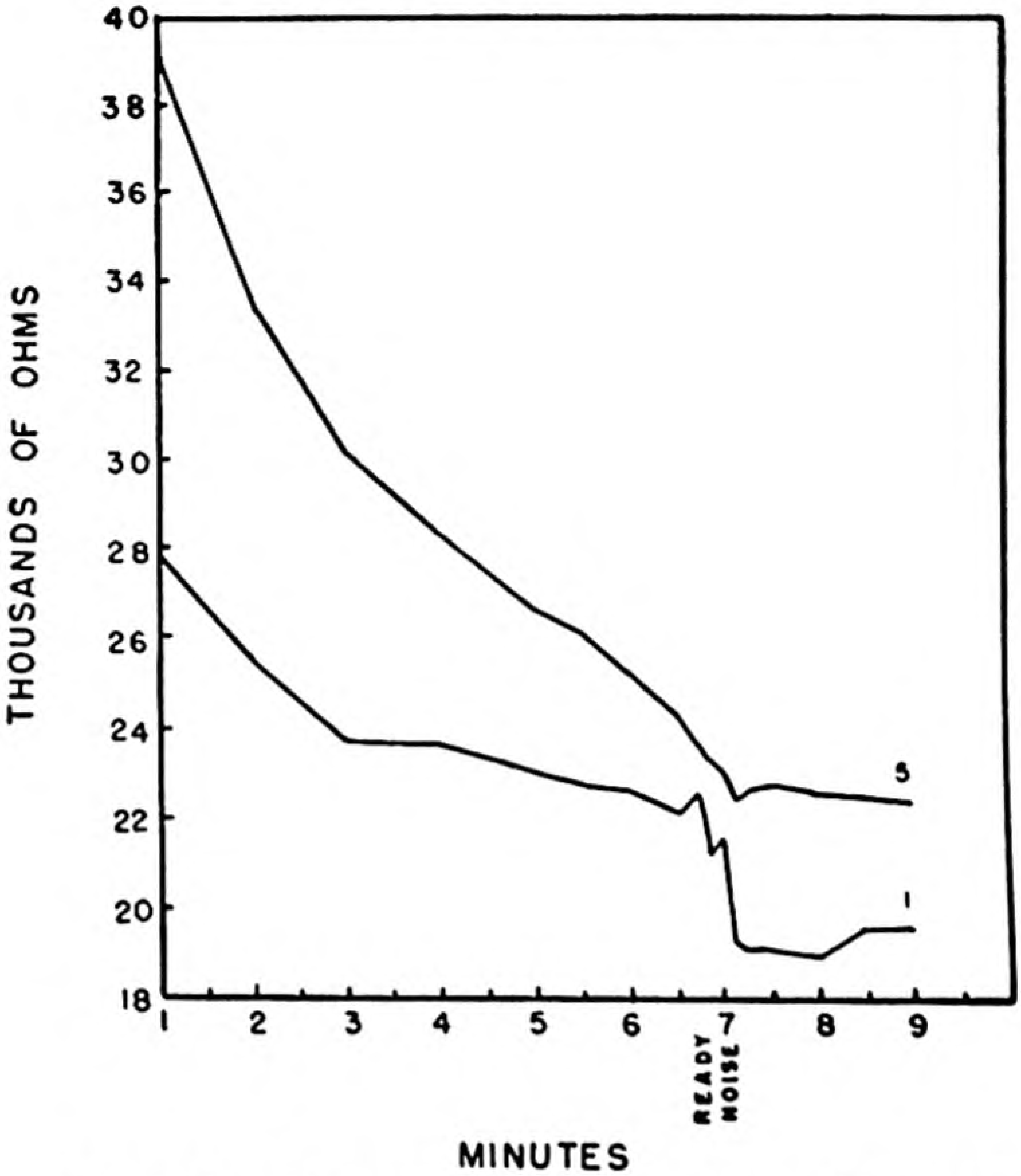


Fig. 110. Changes in Skin Resistance Observed During Experimental Sessions in Which Startle Was Evoked by Means of a Loud Noise. As the session progresses and the subject's expectancy is building up, the skin resistance is lowered and drops further when the "ready" signal and finally the noise itself are given. The difference between the lower curve (day 1 of the experiment) and the upper curve (day 5 of the experiment) shows the effects of adaptation to the situation. (After R. C. Davis, Modification of the galvanic reflex by daily repetition of a stimulus, *J. Exper. Psychol.*, 1934, *17*:510, by permission of the journal and the American Psychological Association.)

signal and the startle caused by the loud noise resulted in a further rapid decline in resistance. That the subject becomes adapted in this situation is indicated by the curves for a subsequent day of

the experiment. Fig. 110, then, illustrates relatively slow changes in the level of skin resistance.

The more rapid change in resistance which can be readily related to the presence of certain stimuli is known as the *galvanic skin response* (GSR). The more rapid change after the "ready" signal and in the presence of the loud sound represents such a GSR (Fig. 110, Day 1). Such deflections may be elicited by a variety of stimuli. Electric shocks and loud noises are uniformly effective. The more general rule is that unexpected stimulation, even though not intense, results in the GSR. The more intense stimuli, however, elicit the GSR with higher frequency and with greater amplitude. The readiness with which the GSR can be elicited by a large variety of stimuli means that it cannot be used to distinguish one emotional situation from another.

What, then, are the conditions for the appearance of the GSR? Physiological investigation has shown that, as typically measured, the appearance of the response indicates increased activity of the sweat glands. The presence of perspiration under the electrodes lowers the resistance to the passage of an electric current. Since the sympathetic division of the autonomic nervous system innervates the sweat glands, the appearance of the GSR indicates that the sympathetic nervous system is aroused, at least to a certain degree. Now, it is known that the sympathetic nervous system tends to act diffusely and is thrown into action by stressful situations. It is, therefore, not surprising to find that the GSR appears in many situations which, on other grounds as well, we would classify as emotional. On the other hand, GSR's have been recorded in many situations which, on other grounds, we would classify as only mildly or not at all emotional.

Facial Expression of Emotions

One of the most commonly observed manifestations of emotion is facial expression. Facial expression derives its special importance from the fact that it is a sign of emotion continually used and reacted to in social relations. Experimental investigations have been addressed to two classes of problems: (1) Are specific facial expressions correlated with specific emotional situations, and (2) how well can individuals judge emotions on the basis of facial expressions?

Correlation of Facial Expression with Specific Emotional Situations. There are two main ways of approaching this problem. Having subjected individuals to a variety of emotional situations, we may correlate their facial expressions either with the type of situation or with the verbally reported emotions. Both approaches have, in fact, been used. It has been found that facial expressions vary too much from individual to individual in a given emotional situation to make it possible to consider any one expression typical of any given situation.

Similarly, it has been found that there is little or no correlation between facial expression and the verbal report of emotion. These findings are based on studies in which individuals were subjected to situations primarily stressful in character. Photographs were taken of their expressions during those situations, and it is on the analysis of these photographs that the above conclusions were based.

It is interesting to note that when subjects were asked to pose, a much better correlation was found between the intended emotion and the posed facial expression. This finding emphasizes that facial expression serves in part as a means of communication. In a given society, a facial expression may serve the same function as a verbal sign to indicate, say, joy, grief, or surprise. Serving as a useful sign, it quickly becomes stereotyped.

Judgment of Facial Expression. Experiments under this heading have dealt with two interrelated problems: (1) How well do judges agree among each other in the name they apply to a given facial expression, and (2) how well do judges agree with the reported emotion or with the intention of the person who poses a given emotional expression?

In a large number of studies, subjects were confronted with pictures of posed facial expressions and asked either to name them or to choose among a list of suggested names. In general, taking the subjects' responses at their face value, little agreement has been found. A wide range of names was used for any given pose. Many experimenters have been prone to conclude that agreement in judging emotions is not much better than chance.

The interpretation of these experiments, however, poses a special methodological problem. In evaluating the names assigned to the

photographs, it is necessary to take into account the different degrees of similarity among the names assigned to any one picture. Thus the names *rage* and *anger* surely denote more agreement than would the two names *rage* and *surprise*. When degree of similarity is thus taken into account, much more agreement among the judges is obtained.

The second experimental question raised in this context concerns the "correctness" of the judgments. If a pose, for example, was intended to depict anger, what proportion of subjects would describe it so? Again taking the names assigned or chosen at their face value, only a small proportion of the subjects are able to make a correct judgment. If, as before, the degree of similarity among the names is taken into account, the percentage of judgments which should be regarded as correct is considerably increased. As Woodworth pointed out, "We need to scrutinize the judgments, asking *how far wrong* they are. We need a scale, rough though it may be, for measuring the error." [1] Thus in one experiment a pose intended to depict *surprise*, was judged as surprise by 52 percent of the subjects. However, if to this percentage are added those who used the names *wonder, astonishment,* and *amazement,* the percentage of presumably correct judgments is increased to 83. Such findings suggest that *posed* facial expressions can be judged with a fairly high degree of accuracy, which is what we would expect in view of their social utility for communication.

As for spontaneous facial expressions recorded in more or less "real" emotional situations, we have already seen that there are no unique patterns of facial reactions for a given situation. It follows that judgments of such expressions cannot be very successful. Experimental investigations have supported this conclusion.

In summary, spontaneous facial expression shows little specificity with respect to either situation or verbal report. Posed facial expressions, on the other hand, constituting a type of language, are much more uniform and are judged by others with a greater degree of success.

[1] R. S. Woodworth, *Experimental psychology,* New York: Henry Holt and Co., 1938, p. 249.

Experiment XXVII

Measurement of Bodily Change in Emotion

A variety of bodily changes are included in what we call emotional behavior. In this experiment, the techniques for the measurement of two types of bodily changes will be illustrated.

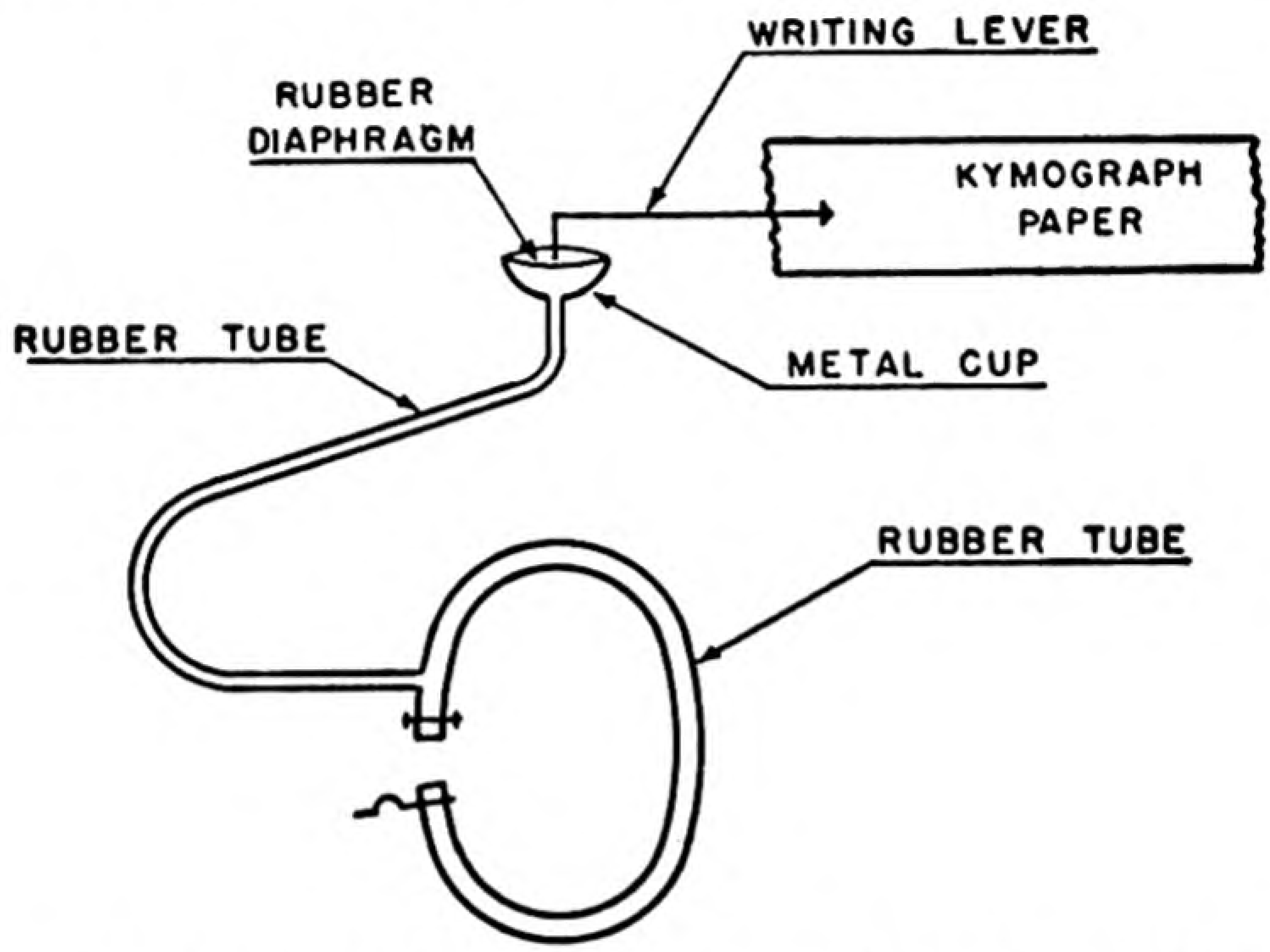


Fig. 111. Schematic Diagram of Apparatus Used in the Measurement of Changes in Respiration.

Purpose. To measure changes in respiration and skin resistance evoked by stimuli intended to produce emotional reactions–startle and fear.

Apparatus. A pneumograph, a tambour, connecting tubes, and kymograph or polygraph are required for the measurement of changes in respiration. The recording system is connected to the pneumograph in the same manner as described above (p. 447) and is represented schematically in Fig. 111. If the paper on which the tambour writes does not move at a constant speed, some provision should be made for a time line.

For measuring changes in skin resistance, a specially calibrated galvanometer, sometimes known as a *psychogalvanometer* or *dermatometer,* is required. This device includes the various parts of the circuit shown in Fig. 109. If a psychogalvanometer or its equivalent is not available, the measurements in this experiment may be restricted to respiratory changes.

To produce startle, a loud noise is a very reliable stimulus. A great variety of instruments may, of course, be used for this purpose. The chief

requirements are sudden onset of the stimulus and high intensity of the sound produced. The effectiveness of the stimulus is increased if it is administered unexpectedly. A loud bell or buzzer may be used conveniently.

To produce fear, an electric shocking device may be used. For the description of a convenient device, see pp. 305 f. This shock is administered to the hand.

Procedure. The subject is seated in a comfortable position. A pneumograph is strapped around his chest, and adjusted until a regular inspiration-expiration cycle is recorded by the tambour. The electrodes of the psychogalvanometer are securely fastened to one hand, and the electrodes for shock to the other.

1. *Startle.* After all preparations have been completed, the experimenter watches the subject's breathing record until a regular rhythm is established. Then, without warning, a loud sound is produced, and at the same time the key for the signal marker is depressed. In this way, it is possible to locate on the record the exact time at which the stimulus was given and also to gauge the approximate latency of the response. To get the most clearcut result, it is advisable to administer the sound as closely as possible to the end of an expiration period. A marked increase in the amplitude of the inspiration is then more likely to appear. This procedure should be repeated several times. The degree of unexpectedness will, of course, be reduced after the first time. To offset the subject's habituation, at least in part, the sound should be given at irregular intervals. The magnitude of the respiratory response, however, will probably decline in the course of the session.

   The unexpected loud stimulus will produce not only respiratory changes but also the GSR. Both changes can be recorded at the same time. Special, and rather complex, apparatus is required for the graphic recording of the GSR. It will be sufficient for most purposes if an additional experimenter reads off the deflections of the galvanometer and records them.

2. *Fear.* To induce the emotion of fear, the experimenter announces that he will administer an electric shock to the subject. This announcement in itself will probably produce changes in both breathing and skin resistance. Following the procedure outlined in Experiment XIX, shocks are administered and gradually increased until a moderate intensity is found. During the procedure of adjusting the intensity of the shock, further changes in breathing and skin resistance will occur which may or may not be measured. To increase the later effective-

ness of the shock, the subject is then misinformed that the intensity will be doubled on the next presentation. This is, of course, not done in fact. Having found a working intensity of shock, the experimenter waits until both breathing and skin resistance are again stabilized. He then unexpectedly administers the shock at the intensity previously determined. As before, the key of the signal marker of the polygraph is depressed simultaneously with the administration of the shock, and the deflection of the psychogalvanometer is read by another experimenter.

As we have emphasized, the inducement of emotional responses in the laboratory is difficult and often is not achieved. The chances of success will be increased if a "naïve" subject is used who is not acquainted with the purpose and the procedure of the experiment. In the part of the experiment using shock, the effectiveness of the stimulus will also depend on the subject's tolerance for shock and his previous exposure to it. Moreover, changes in respiration and skin resistance may occur even though the subjects do not report an emotional response.

Treatment of Results. If the experiment is intended only as a demonstration, it will be sufficient merely to observe the graphic record of respiration and to tabulate the magnitude of galvanometric deflections.

If it is desirable, more thorough quantitative indices may be obtained. The amplitude of inspiration can be plotted as a function of time. The increase in amplitude following the startle or shock stimulus should be apparent in such a plot. If the graphic record permits, the I/E ratio—duration of inspiration over duration of expiration—can be determined and plotted against time. Again the change in the ratio following the critical stimulus would appear in the plot. Finally, it may be interesting to tabulate or plot the changes produced by successive presentations of the stimuli in order to show their decreasing effectiveness as the subject becomes adapted to the situation.

Experiment XXVIII

Judgment of Facial Expression

It will be remembered that posed facial expressions can be judged with a fair degree of success and reliability by a group of subjects. The following experiment demonstrates the typical procedure on which this conclusion has been based.

Purpose. To measure the accuracy of judgment of posed facial expressions and the amount of agreement among the judges.

Materials. A series of photographs depicting various posed facial ex-

pressions is used. In Fig. 112 are reproduced examples from a series of such pictures. Each picture is mounted separately on a cardboard back and identified by a letter or number. The experimenter is provided with a key specifying the emotion intended for each of the pictures. A tally sheet is, of course, necessary.

Procedure. The experimenter presents the pictures in a random order. Each experimenter determines his own order of presentation. The subject is allowed to look at the picture as long as he likes and is instructed to name the emotion which he believes is represented by the picture.

An alternative procedure is to provide the subject with a list of names and ask him to check the most appropriate one. This procedure modifies the experimental situation because it introduces the factor of suggestion. Especially if the judgment is a difficult one, the subject may accept one of the names listed even though he may not consider it an adequate one. To push suggestion to its maximum, it would be possible to attach only one name to each photograph and to require the subject to agree or disagree.

Treatment of Results. The results are used to answer two questions: (1) How correctly do the subjects judge each of the photographs? (2) How well do the judges agree in their use of names?

1. *Correctness of judgment.* For each photograph, the names given by the subjects are listed and the frequency with which each name has been chosen is tabulated. After this preliminary tabulation, the names should be grouped according to their similarity in meaning. Some groups may be reasonably considered as fully correct, some as partially correct, and some as definitely incorrect. The percentage of choices falling into each group is then computed. These percentages provide a quantitative index of the accuracy with which the photographs have been judged. The percentage values will, of course, depend on the leniency or strictness with which the names are grouped as synonymous or closely similar. Such groupings can probably best be made by class discussion. The percentage values for different photographs are then compared to show that some photographs are more accurately judged than others. Finally, a mean index of accuracy for the entire series can be stated, dividing the number of choices classified as correct by the total number of choices.
2. *Extent of agreement among judges.* Extent of agreement among judgments of a given photograph varies in part with accuracy. Thus if all subjects judged a photograph accurately, their judgments would

have to be in 100 percent agreement. There may, however, also be agreement among their inaccurate judgments. Thus 20 percent of the judges may call a picture intended for joy, one of surprise or wonder. Indeed, if the intention of the poser is the principle of classification, there may be 100 percent agreement among the judges on an inaccurate choice. An exact numerical index of agreement is difficult to construct but an adequate, though somewhat rough, measure is pro-

Fig. 112. Stimulus Pictures Used in Experiments on Judgments of Facial Expression. (From J. Frois-Willmann, The judgment of facial expressions, *J. Exper. Psychol.*, 1930, *13*:149, by permission of the journal and the American Psychological Association.)

vided by the number of names or groups of names ascribed to a given photograph. Again, it will be seen that some pictures lead to a greater extent of agreement than others.

Not only are some photographs more difficult to judge than others, but individuals differ in their ability to name accurately the emotion intended by the pose. Some individuals have learned the language of emotional expression better than others. These individual differences can be demonstrated by computing for each judge the number of photographs accurately named by him.

General References

Bard, P., The neuro-humoral basis of emotional reactions. In C. Murchison (ed.), *A handbook of general experimental psychology*, Worcester: Clark University Press, 1934.

Landis, C., The expressions of emotion. In C. Murchison (ed.), *A handbook of general experimental psychology*, Worcester: Clark University Press, 1934.

Landis, C., and Hunt, W. A., *The startle pattern*, New York: Farrar and Rinehart, 1939.

Morgan, C. T., *Physiological psychology*, New York: McGraw-Hill Book Co., 1943, Chap. 17.

Ruckmick, C. A., *The psychology of feeling and emotion*, New York: McGraw-Hill Book Co., 1936.

Woodworth, R. S., *Experimental psychology*, New York: Henry Holt and Co., 1938, Chap. 11.

Young, P. T., *Emotion in man and animal: its nature and relation to attitude and motive*, New York: John Wiley and Sons, 1943.

Measures of Physiological Change

Respiratory Changes

Benussi, V., Die Atmungssymptome der Lüge, *Arch. Ges. Psychol.*, 1914, *31*:244-273.

Blatz, W. E., The cardiac, respiratory and electrical phenomena involved in the emotion of fear, *J. Exper. Psychol.*, 1925, *8*:109-132.

Landis, C., and Gullette, R., Studies of emotional reactions: III. Systolic blood pressure and inspiration-expiration ratios, *J. Comp. Psychol.*, 1925, *5*:221-253.

Totten, E., Oxygen consumption during emotional stimulation, *Comp. Psychol. Monogr.*, 1925, *3*, No. 13.

Blood Pressure

Brown, C. H. and Van Gelder, D., Emotional reactions before examinations: I. Physiological changes, *J. Psychol.*, 1938, *5*:1-9.

Marston, W. M., Sex characteristics of systolic blood-pressure behavior, *J. Exper. Psychol.*, 1923, *6*:387-419.

Scott, J. C., Systolic blood-pressure fluctuations with sex, anger and fear, *J. Comp. Psychol.*, 1930, *10*:97-114.

Skin Resistance

Bayley, N., A study of fear by means of the psychogalvanic technique, *Psychol. Monogr.*, 1928, *38*, No. 176, 1-38.

Davis, R. C., Modification of the galvanic reflex by daily repetition of a stimulus, *J. Exper. Psychol.*, 1934, *17*:504-535.

Haggard, E. A., Experimental studies in affective processes: II. On the quantification and evaluation of 'measured' changes in skin resistance, *J. Exper. Psychol.*, 1945, 35:45-56.

Landis, C., and Hunt, W. A., The conscious correlates of the galvanic skin response, *J. Exper. Psychol.*, 1935, *18*:505-529.

Patterson, E., A qualitative and quantitative study of the emotion of surprise, *Psychol. Monogr.*, 1930, *40*, No. 181, 85-108.

Facial Expression of Emotions

Feleky, A. M., *Feelings and emotions*, New York: Pioneer Publishing Co., 1922.

Landis, C., Studies of emotional reactions: II. General behavior and facial expression, *J. Comp. Psychol.*, 1924, *4*:447-509.

Langfeld, H. S., The judgment of emotion by facial expression, *J. Abn. Soc. Psychol.*, 1918, *13*:172-184.

Munn, N. L., The effect of knowledge of the situation upon judgment of emotions from facial expressions, *J. Abn. Soc. Psychol.*, 1940, *35*:324-338.

Sherman, M., The differentiation of emotional responses in infants: I. Judgments of emotional responses from motion picture views and from actual observation, *J. Comp. Psychol.*, 1927, *7*:265-284.

Sherman, M., The differentiation of emotional responses in infants: II. The ability of observers to judge the emotional characteristics of the crying of infants and of the voice of an adult, *J. Comp. Psychol.*, 1927, *7*:335-351.

»»»» 20 ««««

BEHAVIOR IN SOCIAL SITUATIONS

THE body of this book has been concerned with problems which may well be described as the experimental psychology of the individual. Our main emphasis has been on the individual organism, its sensory and perceptual capacities, the laws governing individual learning, retention, and emotional behavior. But the environment in which man lives is a *social* environment, and social relations with others are important determinants of an individual's behavior. It seems hardly necessary to emphasize or illustrate this point: throughout life, much of our behavior is in one way or another conditioned by our fellow men, stimulated by them and directed toward them. It is to the discovery of the ways in which the general laws of behavior manifest themselves in the interaction of individuals that social psychology addresses itself.

In the broadest sense, any experiment in which a subject and experimenter face each other constitutes a social situation. The motivation of the subject, his acceptance of instructions, his communication and coöperation with the experimeter—all these could be legitimately labeled social variables. In most experiments, however, such influences can be well controlled and do not contribute appreciably to variations in observed behavior. In this chapter, we shall address ourselves explicitly to experiments in which the nature of the social situation significantly determines the behavior of the individual subjects.

The Formation of Norms

Group Influences on Perception and Judgment. In Chapter 11, we emphasized the dependence of judgments on the subjective yardstick of the judge. What is large and what is small varies with the subjective criteria of the judge and depends on the range of

magnitudes to which he is accustomed, whether or not he is aware of the nature of his standards. Social influences—the judgments and norms of others—significantly affect an individual's standards of judgment. The groups of which we are members influence our perceptions and judgments, help to build up the subjective yardsticks which we carry around with us in daily life.

An Experimental Paradigm for Studying the Formation of Norms. It is possible to study experimentally how an individual's standards of judgment are acquired in a social context, how they represent conformity to a social norm. We shall now describe, in some detail, a well-known investigation which has provided us with an experimental paradigm for studying the formation of social norms.[1]

In an experiment designed to study the *formation* of norms in a social context, it is desirable to create a situation in which norms acquired outside the laboratory cannot function. In such a situation, the growth and development of social norms can be investigated. *Autokinetic movement* is a phenomenon well suited for this purpose. When an observer looks at a pinpoint of light in an otherwise completely dark room, the light, though physically stationary, will appear to move after a few moments of fixation (see pp. 213 f.). All observers agree on this phenomenon, although the extent and direction of apparent movement may vary widely from person to person. The absence of a spatial framework for the localization of the light makes the extent of the phenomenon very unstable. Certainly, the observer who first experiences the phenomenon has no well-established standards for judging it.

Although observers enter this situation without standards of judgment, they quickly acquire them. To gauge the influence of a social group on the acquisition of standards, the following experimental design was employed. The subjects were required to estimate the extent of apparent movement. Half the subjects first did so alone (i.e., with only the experimenter present) and then participated in a group situation in which each person announced his estimates aloud. The other half of the subjects first served in a group situation and then worked alone.

[1] M. Sherif, A study of some social factors in perception, *Arch. Psychol.*, 1935, No. 187.

The main findings of the study may now be summarized. First of all, the results clearly indicate that each subject in this highly unstable situation quickly builds up a subjective standard of judgment. Instead of giving an erratic series of estimates, as one might expect in the absence of any physical movement or spatial framework, each individual settles down to a typical range of estimates. Relative to this range (and the ranges may differ considerably from subject to subject), he judges movements as long, short, or medium. How does a social context affect the formation of standards of judgment? Fig. 113 provides the answer. As each individual announces his judgments in the presence of others, the members of the group converge toward a common standard—a social norm is established. In those cases in which the subject first experienced the phenomenon in a group situation, the social norm emerged early and was then carried over into the individual sessions. Those subjects who first established a personal standard and then served in a group converged more slowly toward a common norm but finally did so in all cases. The subjects whose median judgments in successive group sessions are plotted (Fig. 113) in the graphs on the left first served as individuals and then as members of a group. The subjects whose records appear on the right first served as members of a group and then as individuals. Comparison of the two sets of graphs indicates that convergence to a group norm is slowed down by previous experience with the phenomenon.

It is important to emphasize that the group norms did not seem to result from the uncritical acceptance of a standard set by a leader. The subjects announced their judgments in random order and their influence on each other was cumulative until a norm acceptable to all emerged. Sometimes, indeed, the subjects were not aware of the fact that they were being influenced by the judgments of others or even believed themselves free of such influences.

This experiment exemplifies one successful paradigm for the study of social processes in the laboratory. It also drives home an important methodological point: social determinants of behavior may be studied profitably through their effects on such individual cognitive processes as perception, judgment, and thought. The characteristics of the social situation are our independent variables;

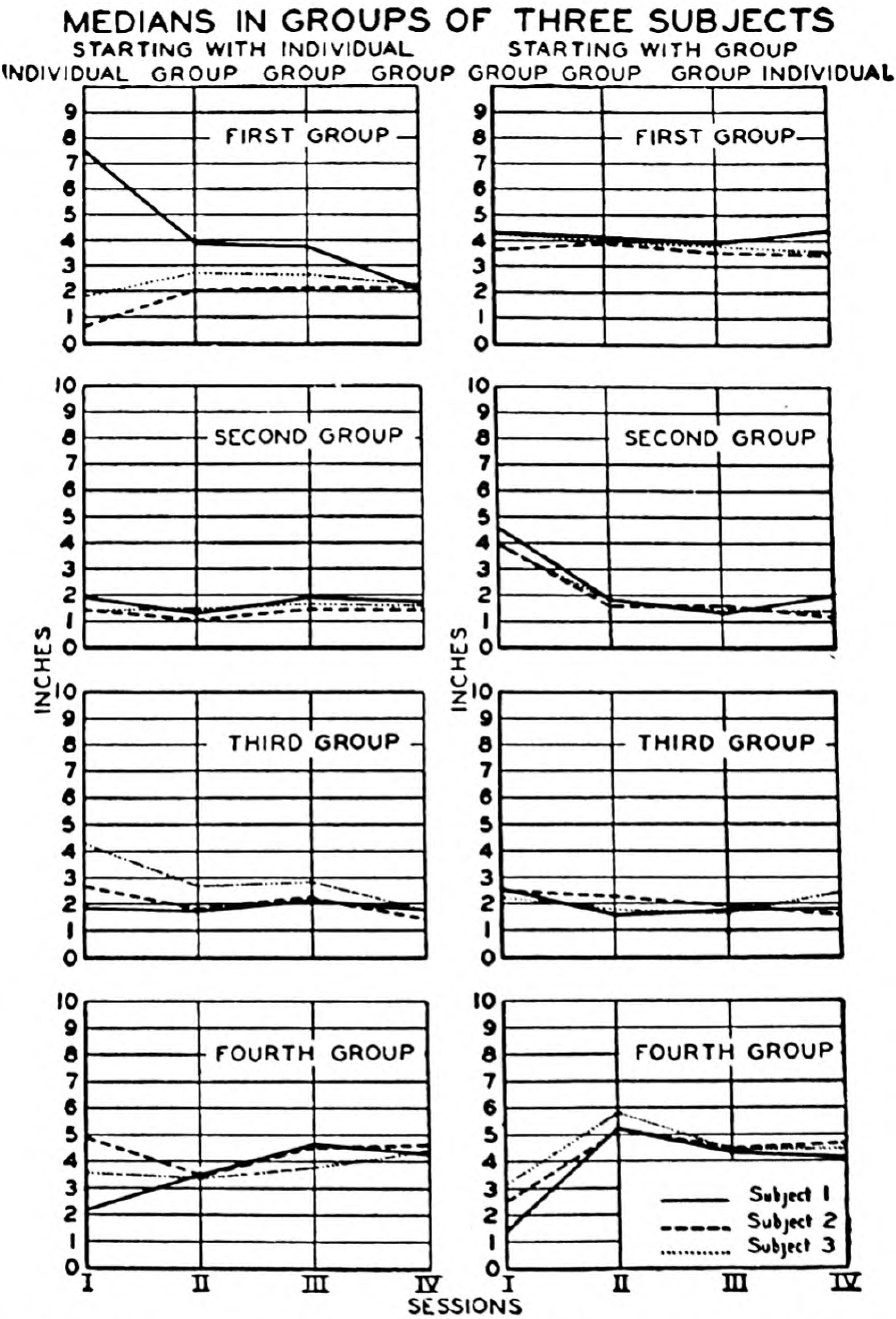


FIG. 113. Convergence of Individual Judgments of Autokinetic Movement Toward a Group Norm. (From M. Sherif, A study of some of the social factors in perception, *Arch. Psychol.*, 1935, No. 187, p. 32, by permission of the journal and the American Psychological Association.)

the observed changes in the individual's behavior, e.g., his perceptual and judging responses, are our dependent variables.

The Factor of Ambiguity. One more methodological point should be considered in connection with this experiment. The situation in which the development of perceptual norms was studied may well be described as "ambiguous." "Ambiguity" refers to the fact that the nature of the stimulus situation does not compel substantial agreement in a sample of observers, the type of agreement that occurs if a group of individuals is required to name the objects of furniture in a room. Autokinesis may be described as ambiguous because the extent of movement observed varies widely from subject to subject. This particular situation was chosen because subjects enter it without any preëstablished standards of judgment. We must not conclude, however, that social norms arise and function only in situations which are "ambiguous" or "unstructured." Even when the situation is made less ambiguous—for example, by requiring subjects to judge actual physical movement and visual extent in a group situation—convergence of individual judgments toward a group norm takes place.

There is good reason, however, why ambiguous situations are frequently used for the demonstration of social influences on perceiving and judging. It is difficult in a brief laboratory experiment to counteract well-established habits of perceiving and judging. The experimenter may choose one of two methods of approach: (1) he may work with ambiguous situations for which there are no well-established norms of judgment, or (2) he may work with unambiguous situations but employ subjects with different attitudes and standards of judgments and demonstrate the effect of such differences on perceptual and judgment behavior. Both types of experiments can be, and have been, conducted to investigate the effects of social variables. The two methods are not, of course, mutually exclusive, nor is ambiguity an all-or-none affair. All gradations of ambiguity and all gradations of differences in subjective attitudes and standards of judgment are encountered in this type of experimental work.

Suggestion and Suggestibility

Suggestion. Our system of social rewards and punishments provides us with strong motivation to comply with current norms, to

accept the standards of the majority and of those in authority even if we had no part in evolving these standards. Often, such conformity becomes almost automatic, eliminating or severely reducing critical evaluation. When stimulus conditions are created which lead to uncritical conformity, we speak of *suggestion.*

What are the features peculiar to suggestion which mark it off from other processes? Many habits are carried out automatically and, in that sense, uncritically, and yet it would be absurd to label all such activities the result of suggestion. The distinguishing characteristics of suggestion must be sought in the social situation in which the behavior takes place. Suggestion involves the manipulation of one individual's behavior by another. The suggester wishes his subject to carry out a certain action. He stimulates his subject—often with the aid of language and other symbols—so that the latter carries out the desired activity. The subject is often unaware of the fact that his behavior has been deliberately manipulated. The suggester utilizes his subject's past conditioning and learning in order to make him conform. Advertising and political propaganda provide many instances of suggestion. Advertisers try to boost their sales not only by demonstrating the merits of their products but also by appealing to extraneous motives—such as "keeping up with the Joneses"—which will intensify the desire to buy. Similarly, politicians often will not rely too heavily on rational arguments in support of their party but will try to tap attitudes and habits which may be unrelated to the issues of the election but will bring out the "right" vote. Whenever such manipulation of behavior is successful and produces conformity, we speak of suggestion.

Types of Suggestion. The subject at whom a suggestion is directed may or may not know that an attempt is being made to manipulate his behavior. On this basis, two types of suggestion may be distinguished: (1) direct suggestion and (2) indirect suggestion.

We speak of *direct suggestion* if the subject knows that the suggester is trying to sway him, that pressure to conform is being brought to bear upon him. He may even realize the nature of the appeal which is being made and be aware of the motives and attitudes which the suggester is trying to arouse. Understanding of the nature of the suggestion does not, however, necessarily lead to successful resistance to the suggestion, provided the motivation aroused

by the suggester's appeal is strong enough. The suggester may make no attempt to conceal his purposes and yet cause his subjects to conform. Political campaigners and advertisers often use direct suggestion with success.

Indirect suggestion occurs when the subject is not aware of the means by which his behavior is being influenced. The suggester succeeds in concealing his aims and methods from the subject. The subject may think that he is acting for his own interest while serving the purposes of the suggester. The advertiser, for example, who argues that to own a musical instrument is a sign of "culture" and thus plays on his reader's aspiration for social recognition, uses indirect suggestion.

Suggestibility. The concept of *suggestibility* is complementary to the concept of suggestion. Whereas suggestion refers to the stimulus conditions which produce conformity, suggestibility denotes an individual's tendency to conform, i.e., to accept the suggestion. In any given situation, wide individual differences in suggestibility may be encountered in a sample of subjects. There will be some who readily and fully accept the suggestion. At the other extreme, there will be those who remain completely unaffected by the suggestion. It may, indeed, happen that the subject will do exactly the opposite of what the suggester wishes him to do. In such cases, we speak of *countersuggestibility*. Between these extremes, there are many intermediate degrees of suggestibility. In a specific situation, degree of suggestibility is a function of many variables. A person who is suggestible in one situation may well resist suggestion in another. There is, indeed, little evidence for a general trait of suggestibility. We turn next to a discussion of the variables of which suggestibility is a function.

The Determinants of Suggestibility

Although examples of suggestion may be drawn readily from daily experience, our greatest advances in understanding the mechanisms mediating suggestion have come from experimental analysis.

The dependent variable in experiments on suggestion is some measure of the subject's conformity, e.g., the frequency and strength of the behavior which the suggester wishes to induce. The inde-

pendent variables are the conditions which the experimenter selects and manipulates in order to render the subject suggestible. The major independent variables in the experimental study of suggestion may be grouped under three general headings: (1) the type of activity subjected to suggestion; (2) the source from which the suggestion emanates; and (3) the characteristics of the subjects at whom the suggestion is directed.

Types of Activity Subjected to Suggestion

Ambiguity and Difficulty of Task. A wide variety of activities has been used in suggestion experiments—cognitive functions, such as perceiving, judging, and problem solving, as well as motor responses, such as movements of the limbs. A classification of experiments according to the different activities subjected to suggestion would be of little purpose, for these activities serve only as vehicles through which the process of suggestion is studied. Instead, it is necessary to isolate those dimensions common to a wide range of activities to which the success of suggestion is lawfully related. Such analysis leads to the generalization that suggestibility varies as a function of the difficulty or "ambiguity" of the task with which the subject is confronted. In general, when a situation is ambiguous, i.e., when there are no well-established standards or procedures for handling it, suggestion has a good chance of succeeding. It is easier to change the amount of autokinetic movement by suggestion than to influence a person's judgments about clearly seen objects in a well-lit room. Similarly, a person trying to perform a very difficult task or to make a very exacting discrimination is more prone to succumb to suggestions regarding methods of solution than one who is handling a problem well within his range of competence and experience.

A Representative Experiment. The generalization relating level of difficulty and ambiguity (and the two are, of course, closely related since what is difficult to perceive or judge is also likely to be ambiguous) with suggestibility is supported by a considerable amount of experimental evidence. We shall limit ourselves here to the detailed description of one study which may well serve as a paradigm. This study was specifically designed to establish and quantify the relation between the difficulty of the problem faced

by a subject and his suggestibility to misleading hints during his attempts to solve the problem.[2]

The materials in this experiment consisted of mathematical problems. These materials were especially useful since they could be graded in difficulty. The subject's readiness for problems of varying difficulty could be roughly gauged by the number of years of mathe-

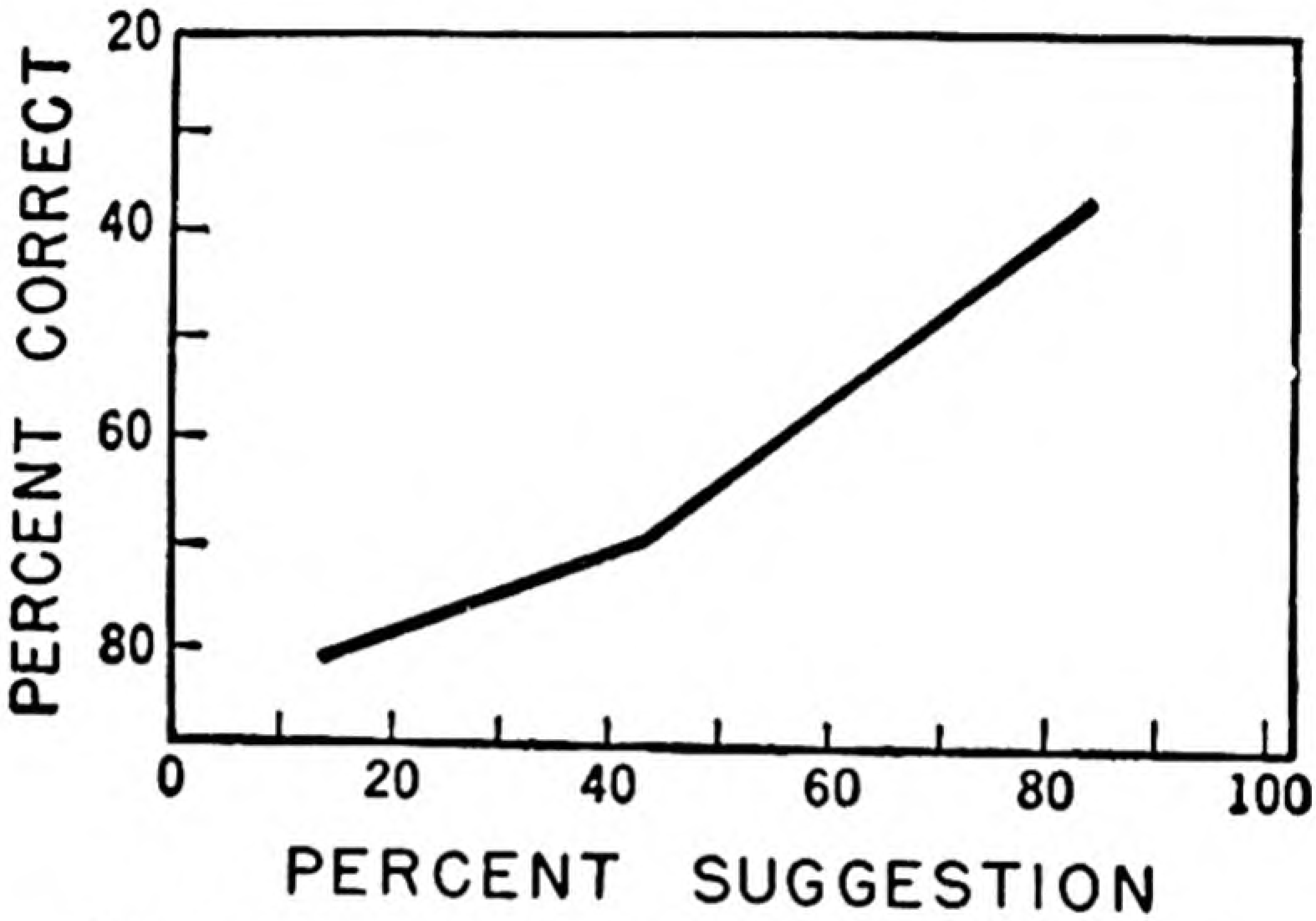


FIG. 114. The Relation of Suggestibility to Difficulty of Task. The easier the task (the larger the percent correct), the smaller the number of successful suggestions. (After T. G. Coffin, Some conditions of suggestion and suggestibility, *Psychol. Monogr.*, 1941, 53, No. 4, by permission of the journal and the American Psychological Association.)

matical training. To conceal the purpose of the experiment, the problems were presented as a "number-facilities test." Suggestions were given in the form of hints pencilled in the spaces provided for the solution of the problems. The subjects were told that the test had been found to be more time consuming than originally expected and that hints were being provided for that reason. Most of the hints were misleading, but in order to establish confidence in them, correct suggestions were given for the first few problems. The test was administered to groups covering a wide range of mathematical train-

[2] T. E. Coffin, Some conditions of suggestion and suggestibility, *Psychol., Monogr.*, 1941, 53, No. 4.

ing. The responses were scored by mathematicians who independently judged each solution as accepting or rejecting the suggestion. The subjects had been instructed not to erase any of their work so that the judges could, in most cases, easily determine whether or not the hints had been accepted.

Level of difficulty was determined independently by presenting the same test, but without any hints, to control groups whose mathematical backgrounds were comparable to those of the experimental subjects. Difficulty was then expressed in terms of the percentage of correct solutions achieved by a control group equated for mathematical training with an experimental group. These difficulty scores were then related to the suggestibility scores. The percentage of solutions attempted on the basis of the hints provided a quantitative index of suggestibility. The results appear in Fig. 114 and clearly indicate that suggestibility increases with the difficulty of the problem, almost in a linear fashion. Similar results are obtained when suggestibility is correlated with amount of mathematical training. The more training a subject had had in mathematics, the less likely was he to utilize misleading hints in attempting to solve the problem.

The experimental procedure which we have just described illustrates an important methodological point. Whenever such variables as "difficulty" or "ambiguity" are used in an experiment, they must be defined in terms of an independent criterion. In this experiment, it will be remembered, difficulty was measured independently of suggestion in terms of the performance of control groups. Similarly, ambiguity can usually be measured, at least roughly, by the extent to which observers agree on the interpretation of a stimulus. Once difficulty or ambiguity has been thus independently defined, it may be used as an independent variable in experimental investigations.

Sources of Suggestion

The success of a suggestion depends on the source from which it emanates, on the nature of the influence which the experimenter brings to bear on his subject. A suggestion is successful to the extent that it eliminates or seriously curtails the subject's disposition to examine a situation critically before he responds to it.

Manipulation of Expectations and Habits. The skillful ex-

perimenter will utilize his subject's habits of response and harness them to his purpose. One way to do so is by manipulating the subject's *set*, by building up certain expectations in him which will prevent him from using effectively his capacity to discriminate and judge. The *progressive line* experiment is an example. A series of lines is exposed one at a time, and the subjects are required to estimate their lengths. In the first part of the series, the lines steadily increase in length, each line exceeding the preceding one by a certain amount. Thereafter, the lines are of uniform length. Nevertheless, increases in perceived length are reported by many subjects (especially children) long after the actual physical increases have stopped. The initial lawful series of increases has established an expectation of continuing increases, an expectation which importantly affects the perceptual judgments. Similar results may be obtained in the judgments of other attributes, such as weight. The number of lines or weights judged larger yields a measure of suggestibility. There are other situations in which the experimenter builds up an expectation in the subject and then "tricks" him. A series of questions may, for example, be arranged in such a way as to induce a set for a certain pattern of answers, e.g., an alternation of *yes* and *no* responses. The subject gets accustomed to this order. If he now encounters a question which requires the same answer as the preceding one, he may fail to respond correctly although he would do so ordinarily. The number of times that he fails to notice the change in the order provides an index of suggestibility.

It is important to note that the source of suggestion in this type of situation cannot properly be called social. The experimenter merely utilizes the fact that perceptual organization and judgment are profoundly influenced by the perceiver's set. Once the set has been built up or reinforced, it readily carries over to situations in which it is no longer appropriate.

There are many instances of apparent suggestibility which are due to the arousal of habitual responses. When a conjurer goes through the motions of throwing a ball into the air and his spectators actually see the ball, they do so because seeing the movement of throwing and seeing the object thrown have been strongly associated in the past. The conjurer arouses this perceptual habit and successfully carries off his trick. When we cry "halt" and our sub-

ject "involuntarily" stops, the success of our suggestion is due to a well-established association between stimulus (the word "halt") and response (stopping).

Prestige Suggestion. In a suggestion situation, one individual attempts to manipulate the behavior of others. Clearly, the attitude of the subject toward the suggester is a critical variable on which the success or failure of the suggestion may largely depend. If the suggester is able to activate his subject's habits of obedience and conformity, if he can restructure and redirect the subject's perceptions and judgments, the suggestion is likely to be effective. It is in the nature of the suggestion situation that an attitude of submission to prestige may become one of the most important conditions of successful suggestion. We speak of *prestige suggestion* if attitudes associated with the acceptance (or rejection) of prestige are primarily responsible for observed changes in behavior.

Prestige is, of course, a general term under which a variety of attitude patterns is subsumed; nor is it always possible to make the nature of the attitudes associated with prestige fully explicit. Prestige is, indeed, not an easy variable to define and quantify in experimental investigations. In many experiments on suggestion concerned with factors other than prestige, prestige factors may nonetheless be an important (and sometimes uncontrolled) variable due to the presence of the experimenter. In the laboratory situation, the experimenter, by virtue of his position of authority, easily becomes a source of prestige suggestion whether or not he wishes to be. Investigations in which children serve as subjects—and children have been widely used in suggestion experiments—the experimenter's prestige may become a very prominent factor.

We shall here consider three questions to which experimental investigations have given at least a partial answer: (1) Is prestige indeed an effective agent of suggestion? (2) What kinds of prestige influences have been found operative and what is their relative effectiveness? (3) How do prestige factors affect the distribution of suggestibility scores in a sample of subjects?

Is prestige an effective agent of suggestion? The experimental evidence definitely answers this question in the affirmative. Experiments on the modification of social and political attitudes, for example, have provided striking evidence. We shall now describe a

typical experimental paradigm for studying the effectiveness of prestige suggestion in bringing about attitude change.

Typically, an experimental group and a control group are employed. A scale measuring attitudes toward some group, issue, or institution is administered to both groups. Some time later, the same scale is again administered, but this time the two groups are treated differently. The members of the experimental group are informed what answers were given to the attitude questions by persons or groups who command prestige in the eyes of the subjects, such as experts in the field under consideration. Or, the statements in the attitude questionnaire are identified as pronouncements of well-known personalities of prestige. The members of the control group, on the other hand, receive the same scale but without any additional information. If the experimental group shows a significantly greater shift in the suggested direction than the control group, we conclude that prestige suggestion has operated. The size of the difference between experimental and control groups provides an index of suggestibility. Studies concerned with attitudes toward a large variety of problems—from prohibition to the literary merits of different writers—have in many, if not in most cases, demonstrated significant shifts attributable to prestige suggestion.

Several methodological problems which arise in connection with this type of experiment should be made explicit. First of all, the experimental and control group should be initially equated with respect to range and intensity of attitudes. Susceptibility to suggestion varies with the intensity of the attitude at which the suggestion is directed. An attitude which is held with great intensity and conviction is less likely to yield to suggestion than one which is accompanied by doubt and uncertainty. For this reason, the results for the experimental and control groups are comparable only if they represent closely similar samples of attitudes.

The control group, of course, serves another purpose: to gauge the shifts in attitude which may be expected to occur, independently of suggestion, between the first and second administration of the attitude scale. Only if the experimental group shows shifts *in excess* of the control group, can the operation of prestige suggestion be inferred. In this connection, the events filling the time interval between the first and second test are of importance. Events outside the

laboratory, e.g., changes in the political or social scene, may influence the attitude of the subjects, either in the same direction as the prestige suggestion or away from it. If such "uncontrolled" events exert a powerful influence, they may override the effects of the prestige suggestion.

Finally, it is important to obtain an independent measure of the prestige of the source of suggestion. After all, what the experimenter considers to be a source of prestige may not be so considered by the subject. Indeed, the experimental findings suggest that subjects often react negatively to the source of prestige and move away from the suggested direction (countersuggestibility). Independent measurements of the degree of prestige of various sources of suggestion must be related to the degree of acceptance or rejection of the suggestions ascribed to these sources.

The paradigm which we have discussed is, of course, applicable to a wider range of problems than the investigation of prestige suggestion. Studies concerned with the modification of attitudes by other techniques, such as various methods of teaching and propaganda, usually employ this experimental design or some modification of it. Nor is this particular kind of experimental procedure by any means the only one for the study of prestige suggestion. Whenever behavior changes are produced by activating the subject's habits of conformity and obedience, prestige suggestion operates.

What kinds of prestige suggestion have been found effective? As we have emphasized, the concept of prestige is an extremely general one, and a variety of attitudes are subsumed under it. Only a few of these attitudes have been isolated and explicitly studied under experimental conditions. Most prominent among these are *conformity to the majority* and *acceptance of expert opinion*. In the experimental modification of attitudes it is usually possible to bring about marked shifts in opinion by indicating the opinions allegedly held by the majority of the group to which the subjects belong or the opinions held by recognized experts in the field. Several investigations have attempted to compare the relative effectiveness of these two sources of prestige. Some studies have given prestige of the majority a slight edge over the prestige of experts but the differences are neither striking nor consistent. If there are true differences in that respect, they are probably dependent on the

specific attitude, the nature of the group, and the status of the particular expert. In any event, no stable generalization about the relative effectiveness of these types of prestige suggestion is possible at this time.

What is the influence of prestige factors on the distribution of suggestibility scores in a sample of subjects? How does prestige suggestion compare in effectiveness with other types of suggestion? We have distinguished between suggestion achieved by manipulation of the subject's habits and expectations, on the one hand, and prestige suggestion, on the other. The distinction between these two types of suggestion is not, of course, a hard and fast one, and in many situations both types of influence are present. Nevertheless, the distinction receives at least partial confirmation from an examination of the relative effectiveness of the two types of suggestion procedures in a sample of subjects.

In Fig. 115, the distributions of suggestibility scores obtained in two representative suggestion situations are compared. In Fig. 115a, the distribution of suggestibility scores on a progressive lines experiment with children is shown (cf. p. 473). The *x*-axis represents composite scores based on the number of times lines are judged larger (in accordance with the suggestion), equal or shorter (in opposition to the suggestion). The larger the score, the greater the subject's suggestibility. On the *y*-axis, the number of cases receiving each of the scores is plotted. Clearly, the distribution of scores is approximately normal in shape. The majority of the subjects are somewhat suggestible, but only a few subjects are completely immune to this type of suggestion and very few show extreme suggestibility (overestimate the lines for a large number of trials). Now, turn to Fig. 115b. Here the results of a prestige suggestion experiment are plotted. The experimenter suggested to his subjects (children) that their hands were getting lighter and lighter and finally would rise in the air. On the *x*-axis of Fig. 115b are plotted suggestibility scores (based on the extent of hand-raising movement the experimenter was able to induce), on the *y*-axis the number of cases receiving each of the scores. This time, with the prestige factor present, the distribution of scores is anything but normal, rather it may be described as "U-shaped," with the majority of cases at the two extremes and a few cases in the middle. Subjects seem either to

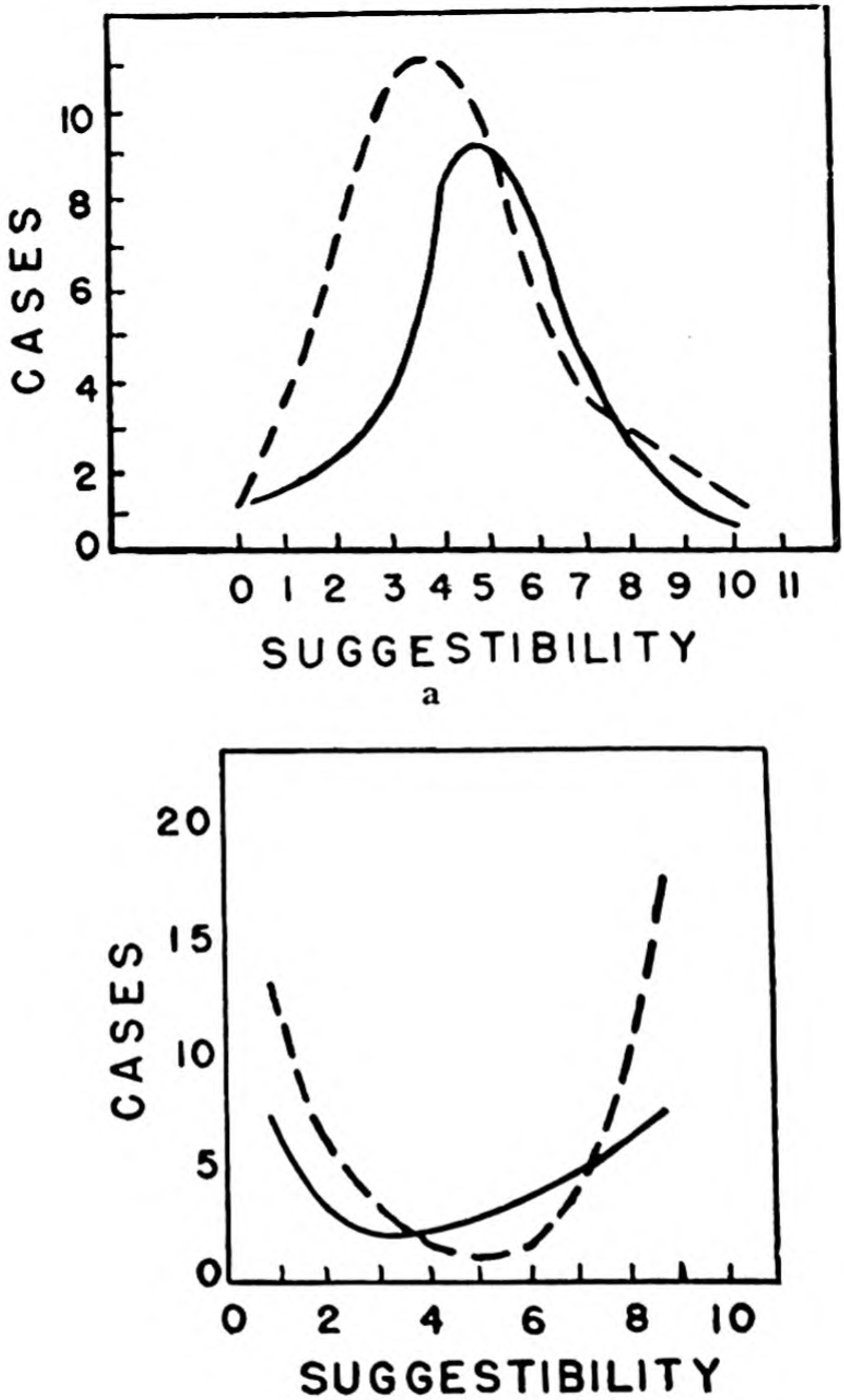


FIG. 115. Distribution of Suggestibility Scores in Two Suggestion Situations. In a are shown the distributions of suggestibility scores obtained from samples of children in a nonprestige suggestion experiment (progressive-lines test); in b the distributions of suggestibility scores in the same samples of children in a prestige suggestion experiment (hand levitation). (From F. Aveling and H. L. Hargreaves, Suggestibility with and without prestige in children, *Brit. J. Psychol.*, 1921, *12*:67, by permission of the journal.)

succumb to the influence of the experimenter or to resist it. Of course, comparisons of distributions such as those shown in Fig. 115 must be made with great caution. The shapes of the distributions depend to a considerable extent on the units of scoring. The units used in Fig. 115a and Fig. 115b are not themselves equivalent, and different distributions might have resulted if different units of measurement had been employed. Nevertheless, there seem to be real differences in the relative frequency with which various degrees of conformity are observed in the two situations.

Who Is Suggestible?

One fact which is common to almost all experiments on suggestion is the wide range of individual differences in suggestibility. A large scatter of suggestibility scores, such as those shown in Fig. 115, is typically obtained. Much effort has been directed at the discovery of those personal characteristics which are correlated with degree of suggestibility. If the personal characteristics associated with high and low suggestibility were fully known, much light could be thrown on the specific psychological mechanisms responsible for the phenomena of suggestion. Although a few such correlations have been established, they have not been extensive enough to contribute significantly to a theory of suggestion. The most stable correlations are between age and sex, on the one hand, and degree of suggestibility, on the other.

Age. Many experimental investigations have shown that children are more suggestible than adults. The suggestibility of very young children cannot be determined readily, since their understanding of instructions may be faulty. There is, however, evidence for increasing suggestibility between the ages of about five and nine, after which susceptibility to suggestion begins to decrease. Children are probably more suggestible than adults because they have fewer well-established habits and standards of judgment which they can bring to bear on the suggestion. For a child, many situations are difficult and ambiguous. Difficulty and ambiguity, as we have seen, are conditions highly favorable to successful suggestion. Moreover, the child has been frequently rewarded for obedience and submission to the adult. This tendency to obey and submit is generalized to the situation in which the suggestion is made.

Sex. Experimental results tend to show greater suggestibility among women than among men. However, the distributions of suggestibility scores of men and women usually overlap to a considerable extent. There is far more variation in suggestibility within each sex than there is between the two sexes. Any explanation of such differences in terms of biological constitution is highly questionable. Differences in education, training, and experience are probably largely responsible for whatever sex differences in suggestibility have been found. For example, subjects in suggestion experiments are often asked to handle some kind of mechanical equipment. Women may have fewer habits and skills to cope with such tasks and, hence, may be more susceptible to misleading suggestions which they cannot evaluate properly. In addition, there may be social pressures at work here. Girls are rewarded for submissive and compliant behavior more often and for a longer time than are boys, and on that basis should be expected to be more suggestible.

Behavior Mechanisms in Suggestion

The term suggestion does not refer to any unitary process of behavior. Suggestion is a purely descriptive term and denotes a variety of behavior mechanisms which make it possible for one individual to manipulate the behavior of others. It is not possible to enumerate all the habits and tendencies which may operate in a suggestion situation. Certainly past learning and conditioning play an important part. When we point at an imaginary airplane in the sky and somebody thinks he sees it quite vividly, we say he is "suggestible." It would be more accurate to say that the stimuli of pointing and pronouncing the word "airplane" evoke a well-established perceptual habit—seeing the airplane. Similarly, in the progressive lines experiment, success of the suggestion is due to the operation of a habit. The association between the appearance of the lines and the seeing of an increase in magnitude continues to function even when the lines have become physically equal. The behavior mechanisms operating in suggestion, then, are special instances of learning and conditioning. It is the suggester's task to provide the stimuli that will activate the learned response which he desires to occur.

The mechanisms operating in prestige suggestion should ultimately lend themselves to a similar analysis. The presence of an

individual of prestige or the citation of an expert opinion throws into action patterns of responses which were reinforced in the past in similar situations. Such patterns of responses may include perceptual reorganization of the situation and changes in verbal responses. The emotions which the presence of the individual of prestige arouses probably serve to facilitate and strengthen the activation of such patterns of response. The point which we should like to emphasize here is that in studying suggestion, we investigate the results of past learning and conditioning as they manifest themselves in a special type of social situation.

Work in a Group Situation

Our discussion in this chapter is focused on the experimental analysis of behavior *as it is determined by the actions of others.* We have considered the formation of norms in a group situation and some conditions under which conformity behavior is evoked. In this section, we shall be concerned with the ways in which an individual's work and productivity are affected by the presence of others and their actions. The influence of a group on individual performance will be discussed in reference to four types of experimental situations: (1) the effect of an audience; (2) the effect of the presence of coworkers; (3) work in a competitive situation; (4) the results of group discussion and coöperation.

Audience Effects

Typical Experimental Procedures. In experimental investigations of the effects of an audience on the performance of an individual, a control condition and an experimental condition are usually employed. In the control situation, the subjects work alone on some task, such as mental multiplication or a test of motor dexterity. In the experimental situation, they perform the same task in the presence of an audience. Their performances under the two conditions are compared and differences between them are ascribed to the effect of the audience.

This experimental design is adequate provided the two conditions have been equated in all respects other than the presence of an audience. One of the most important precautions is to equate the subjects' level of practice in the two situations. If the subjects enter

the two situations with different amounts of training, differences in performance will be due, to an indeterminate extent, to differences in training, and the effect of the audience cannot be evaluated adequately. To equalize practice level, one of two procedures may be followed: (1) if the same subjects serve in both situations, they should be given a task which they have thoroughly mastered and in which no further improvement may be expected; (2) if different subjects are employed in the two situations, the controls and experimentals should first be equated in training. While the experimental subjects work in front of an audience, the control subjects continue to work alone. Thus both groups start from the same level of proficiency and have an equal chance to benefit from further practice. If they fail to do so, the effect of the audience is probably responsible for the discrepancies in performance.

The terms "alone" and "audience" need, of course, to be defined clearly. In some experiments, performance of subjects working in the presence of the experimenter is contrasted with their performance in front of a larger group. In other experiments, the subject is all alone in one situation and works in front of a group in the other. Obviously, the experimenter is an "audience," and sometimes an exacting audience with considerable prestige. Experiments using different definitions of the "alone" situation cannot be expected to yield comparable results.

"Audiences" may also vary in many ways. First, they may differ in size. A variety of relationships may exist between the subject and the members of the audience. The audience may consist of friends or strangers, of the subject's peers or of his superiors in rank. The members of the audience may be passive in their attitude toward the subject, they may actively encourage and support him, or they may be hostile. Clearly, each of these characteristics may radically affect the influence of the audience on the subject's performance. Not only the attitude of the audience toward the subject, but also the subject's attitude toward the audience is of importance. Much will depend on how much experience the subject has had in public performance and how much he values the approval of the audience.

Experimental Results. Considering the many ways in which the nature of an audience can vary, it is not surprising that there are no clear-cut generalizations about the effect of an audience on an

individual's performance. In comparisons of work in an "alone" situation and in front of an audience, both facilitating and inhibitory effects have been reported. As far as a trend can be distinguished, it appears that a passive audience is not conducive to the best possible performance. Although there are many exceptions to this rule, even the exceptions are not clearly in the direction of better performance in front of an audience. Thus it was found in some experiments that the presence of spectators spurred the subjects on to greater speed but at the expense of quality and accuracy of performance. The presence of an audience may also render subjects more cautious and circumspect, as, for example, in the reproduction of memory material in front of a group of spectators (see p. 415). As to active audiences, their effect varies, of course, with the nature of their activity, whether it is encouraging or discouraging.

Since so much depends on the individual subject's reaction to the presence of an audience, it is not surprising to find that experiments on audience effects usually show a considerable amount of individual differences. In the presence of the same audience, some subjects improve while others lose in efficiency. For this reason, the *averaging* of performance scores in such experiments may become meaningless. Important, but opposite, effects on individual subjects may cancel out. Hence, many experimenters have used statistical indices other than measures of central tendency to evaluate the effects of an audience. Thus the percentage of subjects showing an improvement and the percentage showing decrements can be compared. It is also possible to measure the absolute amount of shift from one condition of work to the other (i.e., disregarding the direction of the shift) in order to answer the question of whether or not the audience has any significant effect at all.

Presence of Co-workers

Typical Experimental Procedures. Much of our daily work—in the factory, in the office, and in school—is done in the presence of others. Co-workers, then, are an important feature of the stimulus conditions under which many tasks are performed. Experimental investigations have been directed at the question of whether an individual performs differently (better or worse) in the presence of co-workers than he does alone. Again, the typical experimental de-

sign calls for a comparison of performances under two conditions: working alone and working in the presence of others who are engaged in the same task.

The controls and precautions which we have discussed in connection with experiments on audience effects must be observed again. Valid comparisons between the two work situations can be made only if such factors as training and familiarity have been equated. If the same subjects serve in both situations, it may be advisable to use a counterbalanced design: half the subjects serve first alone and then in the group situation, with the sequence of conditions reversed for the other half of the subjects.

In this section, when we speak of the effects of co-workers, we mean the influence which the sheer physical presence of others engaged in the same task has on the performance of an individual. The sight and sound of others, their movements and activities, are stimulus conditions the importance of which for individual performance and productivity can be experimentally determined. It is exceedingly difficult, however, to isolate the effects of the sheer physical presence of others. When individuals work alongside each other, important motivational factors come into play: attitudes of competition and coöperation arise, some individuals become leaders who set the pace for the group, and so on. Experimenters interested in studying the facilitating or inhibiting influence of the sheer presence of co-workers have usually tried to minimize motivational changes by removing all incentives for rivalry. In such experiments, subjects are explicitly warned against assuming a competitive attitude, the results of their work are not announced, and care is taken to prevent one subject from finishing before the others. By such measures, the factor of rivalry can, indeed, be reduced in importance, but it is doubtful that it can ever be completely eliminated.

Experimental Results. How does the presence of co-workers affect the performance of an individual? First of all, there is good evidence that people work at *greater speed* in a group situation than they do alone. Increases in output have been shown with a variety of tasks, such as multiplication, word association, cancellation of letters, etc. The presence of others increases the sheer amount produced—a phenomenon described as "social facilitation." This

result is fairly typical but by no means universal. Some slow down rather than speed up in a group situation. In general, those initially slow show the greatest increase in speed, whereas subjects who are fast workers to begin with work only a little faster in the group situation. The latter may be working close to the limit of their capacity to begin with. It is not always clear to what extent this differential effect on slow and fast workers is a "regression" phenomenon.

Along with increases in speed, work in a group situation may often result in poorer quality of performance. More problems may be attempted by a subject in a group than when he works alone, but he may actually do fewer of them correctly. In the presence of co-workers, his performance will typically be more variable. As far as quality and accuracy are concerned, advantage seems to be on the side of solitary performance. In a group situation, there are more potential distractions, and the sustained attention necessary for high quality performance is more difficult to achieve than in the alone situation. The greater the demands which a task makes on logical and systematic thinking, the greater may be the decrement in quality caused by the presence of co-workers.

It is unlikely that it is ever possible to eliminate motivational factors, such as rivalry, entirely, even when the subjects are assured that there are to be no comparisons among individual performances. The following experiment illustrates the subtle ways in which rivalry may operate in group work. One investigator used two alone situations and a group situation.[3] In one of the alone situations, each subject worked in a room by himself but knew that there were subjects in other rooms simultaneously working on the same task. In the second alone situation, the subjects came to the laboratory at different times and worked completely alone. The two alone situations were then compared with the group work condition, in regard to speed and accuracy of performance. The first of the alone situations, in which the subject knew that others were working simultaneously on the same task in different rooms, and the group situation gave very similar results: greater speed and less accuracy than in the completely solitary work situation. Clearly, analysis of

[3] J. F. Dashiell, An experimental analysis of some group effects, *J. Abn. Soc. Psychol.*, 1930, 25: 190-199.

group effects in terms of the sheer physical stimuli provided by the presence of others must remain inadequate. The analysis of the motivational factors which the presence of co-workers brings into play is indispensable for an understanding of the determinants of performance in a group situation. To the ubiquitous problem of competition we turn next.

Competitive Work Situations

Typical Experimental Procedures. The experimenter who wishes to study the effects of competition upon performance faces several methodological problems. First of all, he must find ways to vary systematically his independent variable—degree of competition. How can degree of competition be varied experimentally? Carefully worded instructions provide a first method. By such instructions, competitive attitudes may be considerably intensified, especially if they are given by a person in authority, such as a teacher in a schoolroom situation. The instructions can be supported by various techniques: prominent publication of the scores achieved by the competing individuals or groups, pointed comparisons of achievements, commendations to the successful competitors, and exhortations to those lagging behind. Finally, prizes may be offered to the winners as additional incentives to compete.

Having created a competitive work situation, the experimenter needs, of course, a noncompetitive work situation for purposes of comparison. It is probably more difficult to create in the laboratory a truly noncompetitive work situation than it is to create a competitive one. Many experimenters perforce rely heavily on instructions to their subjects to eliminate or reduce the motivation to compete. Competitiveness is also mitigated by avoiding publication of results and any comparisons among individuals. Nevertheless, some indeterminate degree of competitiveness is probably present in all group situations. Many experiments, moreover, have been carried out in classrooms where competition for grades and recognition is so much part of the psychological climate that even elaborate instructions will probably fall short of eliminating rivalry. To summarize: in experiments on group work, the experimenter may with some success vary the intensity of competitive attitudes. In a competitive social environment such as ours, it is doubtful that it is

possible to create in the laboratory a truly noncompetitive group work situation.

Factors other than degree of competition should be equalized for the various experimental groups. Both the competing and noncompeting groups should perform the same task, or a task equated for difficulty. Competing and noncompeting subjects should be matched for amount of training and ability. When all these controls have been carefully observed, the performances under competitive and noncompetitive conditions can be compared properly. Significant differences among the various groups may then be ascribed to the effects of competition.

Experimental Results. There is general agreement that intensified competitiveness leads to greater work productivity. The results of a typical experiment may be cited to illustrate this general finding.

The subjects in this experiment were grade-school children.[4] (Children are frequently used in experiments on competition since they respond more readily to instructions and incentives.) Their task was the solution of arithmetic problems. An experimental group and a control group were formed, equated for age, sex, and ability. The members of the control group performed the tasks as part of their regular school work. The experimental subjects were subdivided into two equated groups and urged to compete against each other. The experimenter reinforced the competitive attitude by publicly announcing the names of the winners and writing their names on the blackboard. The losers were admonished to increase their efforts. This procedure was continued for five days. In a schoolroom situation, these methods of intensifying rivalry proved most effective. The output of the competing subjects greatly exceeded that of the controls. On the final day of the experiment, there was a difference of 41 percent between the two groups. The competing subjects were superior not only in output but also in accuracy of solutions. This experiment may well serve as a paradigm for investigations of work under competition. The procedures and findings are representative of many other studies.

The general finding, then, is that intensification of a competitive attitude increases productivity without necessarily injuring the

[4] E. B. Hurlock, The use of rivalry as an incentive, *J. Abn. Soc. Psychol.*, 1927, 22:278-290.

quality of performance. The *degree* to which a competitive attitude affects the level of work depends, of course, on a number of parameters. A few of these variables will be briefly considered next.

Individual vs. group success. There may be competition among groups and among individuals. Class may be pitted against class or student against student. Several experimental studies have compared the effectiveness of competition for individual rewards and competition for rewards given to the group, of which the individual workers are members. The experimental results indicate that individuals work harder and better when competing for individual reward and recognition than when the success of a group is at stake. The differential effects of individual and group incentives become especially pronounced if the experiment is carried on over a protracted period of time. Those working for individual success sustain their effort longer and show less variability of performance than those working for group success. An independent check on the greater incentive value of individual rewards may be obtained by permitting workers to divide their time freely between work for individual and group reward. Subjects choose to devote the bulk of their time to efforts in their own behalf and work more effectively while working for themselves than when working for the group. Such are the experimental findings. To what extent such "egotistic" tendencies are modifiable by training is another question.

Nature of the task. What effect the creation of a competitive atmosphere will have depends in part on the nature of the task in which the subjects are engaged. If the task is in itself interesting and pleasant, subjects may work close to the top of their capacity and the additional incentive of competition can have only very limited effects. On the other hand, if the task is dull and the subject is poorly motivated, the introduction of competition may boost the output level considerably. The available experimental data have tended to confirm these expectations.

Individual differences. Some subjects are more responsive to the introduction of competitive conditions than are others. As in the studies of social facilitation, those whose work is initially poor, show greater improvement under competition than those whose work is good to begin with. When the performance level is already high, there is, as it were, little room for improvement. Furthermore,

those whose initial performance is inferior are often poorly motivated, and the introduction of competitive conditions provides them with the incentive to work on a level representative of their ability.[5]

Coöperative Work Groups

Coöperation and competition are twin motives sustaining and regulating the aspirations and activities of an individual in a social context. In many of our daily enterprises, both motives are operative. In experimental studies, we usually try to study each of these motives individually, to create situations which are either competitive or coöperative. In this section we are concerned with the analysis of coöperative work.

Typical Experimental Procedures. Several important experimental investigations have focused on the question of whether the achievements of a coöperative group *qua* group are superior to the achievements of the individual members working by themselves. Do the opportunities for interaction and coöperation afforded by a group situation add to the combined productivity of the individual members?

Experimental studies of this question call for the familiar comparison of experimental and control groups. In the control condition, each individual attacks a task by himself, such as solving a puzzle, arranging words into sentences, drawing a conclusion from a set of facts, judging a picture, etc.[6] In the experimental condition, a committee collaborates on the problem and presents the results of its common labor to the experimenter.

This general experimental scheme needs to be filled in by choices of specific tasks and specific coöperating committees. Clearly, tasks will vary in their susceptibility to group treatment. There are some tasks which do not allow much group discussion or coöperation and in which the products of the individual members are simply added together. Thus when a group is given the job of finding synonyms or antonyms for a series of words, each group member will find a certain number of words and the activity of the other members is

[5] As pointed out above, the factor of regression must be carefully considered when relating improvement to initial performance level.

[6] In this discussion, we shall not be concerned with the problem of increasing efficiency and productivity by division of labor and assembly-line methods. We shall limit ourselves to problem-solving behavior in coöperative group situations.

limited to the checking of occasional mistakes. In addition, the presence of co-workers may produce "social facilitation" (see pp. 483 f.). On the whole, however, such tasks do not lend themselves to significant analysis of coöperative work. On the other hand, there are tasks, the solution of which involves a series of choices and decisions, in which the give and take of group decision may radically affect the course of solution and its eventual success. Analysis of problem solving in coöperative groups may help to throw light on those variables which determine the behavior of committees, juries, administrative boards, and similar collaborating groups.

The size and personnel of committees may vary in an almost bewildering number of ways. The more specifically the experimenter has formulated his problem, the less haphazard will the choice of the committee be. If we are interested in studying the leader-follower relations which develop in coöperating groups, we may want to choose members of unequal ability in order to facilitate the emergence of a leader. On the other hand, if our main concern is with studying the critical function performed by the group, we do well to choose members of about equal skill and competence to maximize the opportunities for interchange of ideas and criticism. The optimal size of the group will depend on how intensively we wish to study the behavior of each individual member. Moreover, the nature of the problem to be solved will impose a limit on the number who can effectively coöperate in its solution.

In comparing individual with coöperative performance, we must, of course, equate difficulty of tasks and ability of subjects in the two situations. Counterbalanced designs, in which the same subjects serve in both situations, may be employed. Half the subjects should work first individually and then in a group; the other half follow the reverse sequence.

The significance and fruitfulness of such an experiment may depend in large measure on the experimenter's choice of specific categories of behavior to be observed and recorded during the experiment. A mere tabulation of the number of tasks solved by individuals working alone and in a group does not do justice to the problem. The ways in which group discussion develops, and the type of interaction among members are often the most revealing results of the experiment. Careful observation of the group at work

and detailed records of group discussion and work procedures are of the essence. These records can then be ordered into such categories as "number of suggestions made by each member," "number of correct and incorrect suggestions accepted and rejected," etc. In this manner, the particular types of work procedures characteristic of a coöperating group can be described and analyzed.

Experimental Results. There is no question, on the basis of the experimental evidence, that coöperative groups are more effective and productive than individuals. As we have emphasized above, the superiority of the group is not always due to the operation of the same factors. Sometimes the high performance of a group represents merely the summation of individual performances; in other cases, it is the opportunity for discussion and correction of mistakes which is mainly responsible for the superiority of the group.

Different tasks benefit to varying degrees from group efforts. In one investigation, a series of different tasks were solved by subjects working individually and in groups. In the accompanying table, these tasks are listed and ranked according to the degree of superiority of the group performance over individual performance.[7]

Rank Order of Tasks According to Degree of Group Superiority
(Total achievement, including quality)

Task	Rank
Finding words meaning opposite of given words	1
Solving a cipher	2
Drawing conclusions from given facts	3
Sentence completion	4
Listing steps to solve a problem	5
Composing limerick	6
Comparison of numbers	7
Reading comprehension	8
Original intelligence problem	9

Both types of group influence mentioned before may be found in this table. The highest rank is occupied by a task (finding

[7] From G. Murphy, L. B. Murphy, and T. M. Newcomb, *Experimental social psychology*, New York: Harper & Brothers, 1937. After G. B. Watson, Do groups think more efficiently than individuals? *J. Abn. Soc. Psychol.*, 1930, *21*:79-109.

antonyms) in which the group results represent the addition of individual achievements. In the second and third place, we find tasks the solution of which by a group of workers is advanced by public discussion, acceptance and rejection of methods of solution.

Careful analysis of the procedures by which a coöperating group solves a problem has shown that one of the most important functions of the group is the elimination of errors and the rejection of false suggestions for solution. Statistical analysis of the trends of group discussion has shown that incorrect suggestions are much more likely to be rejected by one or the other member of the group than are correct suggestions. Moreover, erroneous suggestions are much more frequently rejected by other members of the group rather than by their originators. As a result, group superiority is especially pronounced when the solution of a problem depends on a series of steps or choices, each of which can be considered, discussed, and voted upon by the members. A group is likely to solve more such problems than an individual. If the group fails to solve the problem, the error responsible for the failure will probably occur later in the course of the solution than is typically the case with individuals working alone. The early errors to which an individual may fall victim have a good chance of being checked by group discussion.

Experiment XXIX

The Formation of Norms[8]

Purpose. To study the formation of norms in judgments of autokinetic movement.

Apparatus. The diagram shown in Fig. 116 shows the apparatus to be used in this experiment. For a full explanation, see the legend of the figure.

Experimental Design. The basic design of this experiment calls for the use of two groups of subjects. The members of Group A first serve individually and then in groups of three. The members of Group B first make their observations in groups and then individually, thus:

Group A: Individual sessions followed by group sessions
Group B: Group sessions followed by individual sessions

The time interval between sessions should be the same for Group A and Group B.

[8] After M. Sherif, A study of some social factors in perception, *Arch. Psychol.*, 1935, No. 187.

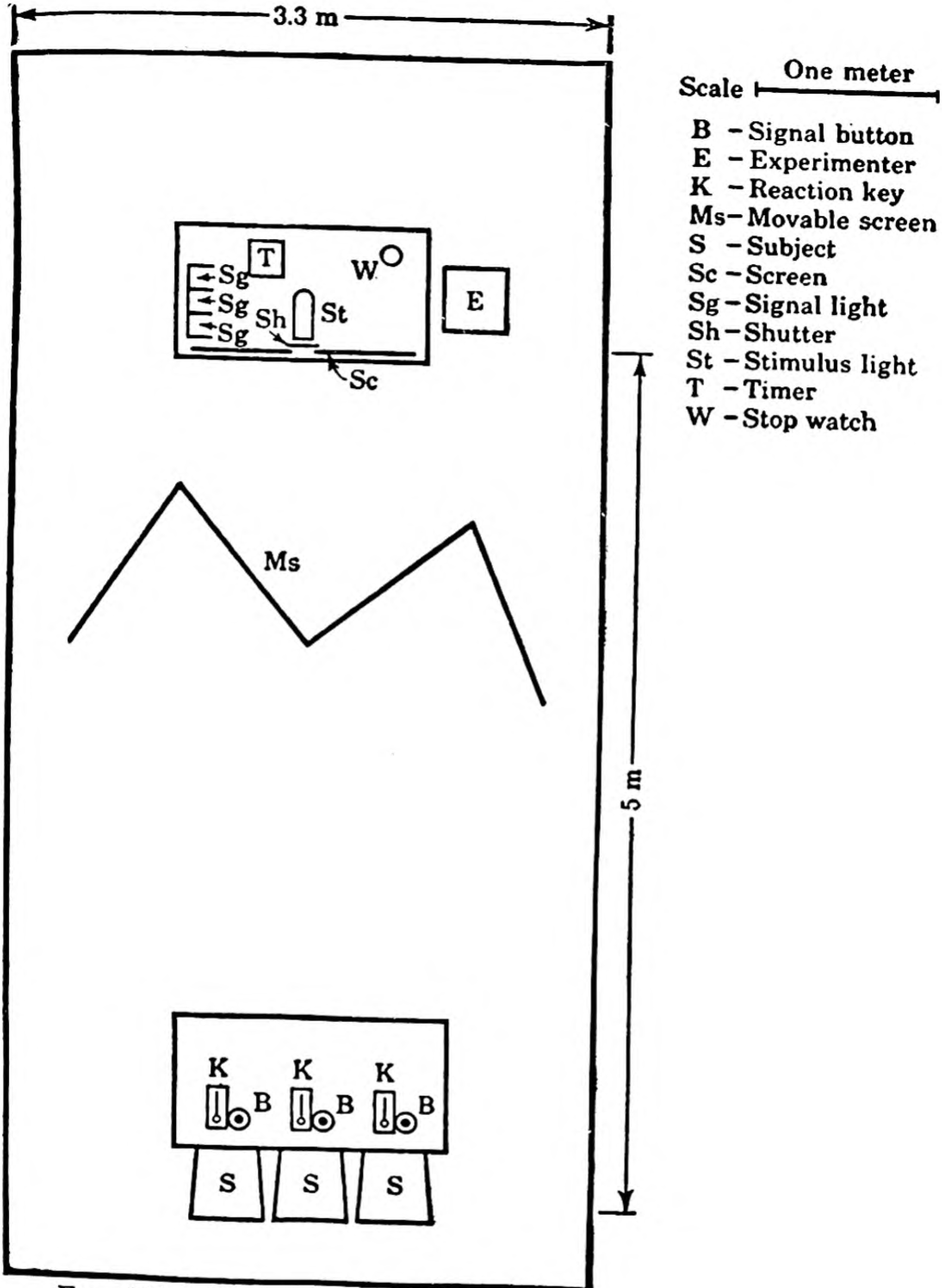


FIG. 116. Experimental Setup for Experiments in Autokinetic Movement. (From M. Sherif, *An outline of social psychology*, 1948, p. 165, by permission of Harper & Brothers.)

Care should be exercised in selecting the members of the groups. If possible, the nature of the relationship among the members should not vary too much from group to group. It is inadvisable to have one group consisting of close friends, the other of strangers.

Procedure: Individual Sessions. The subject is blindfolded, led into the dark room and seated in his chair. The experimenter reads the following instructions to the subject:

> "When the room is completely dark, I shall give you the signal READY and then show you a point of light. After a short time, the light will begin to move. As soon as you see it move, press the key. A few seconds later, the light will disappear. Then tell me the distance it moved in inches or in fractions of an inch. Make your estimates as accurate as possible."

The experimenter removes the blindfold and allows the subject some time to dark-adapt. He then gives the first "ready" signal and switches on the pinpoint of light. As soon as the subject has pressed his key to indicate that movement has begun, the experimenter begins to time an interval of 2 seconds. At the end of the 2-second interval, the light is switched off, and the subject makes his estimate of the extent of apparent movement. This procedure is repeated for twenty trials. (A larger number of trials is desirable but may prove too time-consuming if the whole experimental design is to be carried out in the course of one laboratory session.) At the end of the first series of trials, the subject is allowed a brief rest interval in the dark room, and, thereafter, another series of twenty (or more) judgments is obtained. All the estimates are recorded by the experimenter. Since the room is totally dark, the experimenter will do well to note each estimate on a separate sheet of paper.

Procedure: Group Sessions. The same basic procedure is followed as in the individual session except, of course, for the simultaneous presence of three subjects. The subjects are instructed to announce their judgments in any order they wish. The subjects are encouraged to change the order in which they give their judgments from trial to trial instead of adopting a set order. Whenever a subject announces his estimate, he presses a signal button which activates a light on the experimenter's table. In this manner, the experimenter can identify the subject without relying on recognition of his voice.

Treatment of Results: Group A. The members of this group first make their observations individually and then in groups of three. We examine the data with a view to answering two questions: (1) Did the judgments which the subject made individually tend to become increasingly stable, i.e., did the subject establish a range within which his

responses tended to fall and a central tendency around which the judgments clustered? (2) Did the members of the group, each of whom had had some experience with the autokinetic phenomenon, converge upon a common norm?

To answer the question of whether an individual's judgments in this unfamiliar and "ambiguous" situation tend to stabilize within a limited range and around a central value, we examine the two series of twenty individual observations. First, we divide the judgments into successive blocks of five—a total of eight blocks. For each of these eight blocks, we compute the median and the range. We may determine by inspection whether the ranges tend to stabilize. We then plot the medians of successive blocks of five judgments. If the subject has, indeed, established a personal norm, the curve obtained by plotting successive medians should level off as the experiment proceeds.

Did the individuals in the group converge toward a common norm? For the group situation, we again divide the judgments into successive blocks of five, and for each block of five, determine the median and the range. We may find out by inspection whether or not the ranges of the three members become more and more alike in the course of the session. In addition, we plot the medians of the three subjects' judgments for successive blocks of five. If a common norm is established, the medians of the three subjects' judgments should converge, i.e., there should be a "funneling effect," such as that shown in Fig. 113. It is, of course, desirable to apply tests of statistical significance to the shifts in the medians.

Treatment of Results: Group B. These subjects start with observations in a group and then give their judgments individually. There are again two questions which the data may answer: (1) Did the members of the group converge toward a common norm? (2) Did the norm established in the group carry over into the individual sessions?

The question of whether the subjects in the group converged toward a common norm can be answered in the same manner as was suggested for Group A, i.e., in terms of medians and ranges for successive blocks of trials.

Did the group norm, if such was established, carry over into the individual sessions? To answer this question, we determine (a) whether the judgments in the individual sessions tend to fall within the same range as in the group session, and (b) whether there is a significant shift in the median judgments from group to individual session. A fair answer to these questions can be obtained by comparing the judgments in the *last* block of judgments in the group situation with successive blocks of

judgments in the individual situation. To represent the results graphically, we again plot the medians for successive blocks of judgments.

Comparisons of Group A and Group B. Comparison of the results for Group A and Group B will yield the answer to an interesting question: Do the members of Group B converge more readily toward a common norm than the members of Group A? The members of Group B, it will be recalled, have their first experience with autokinetic movement in a group situation, whereas the members of Group A may enter the group situation with personal standards of judgment already established. We should expect, therefore, convergence toward a group norm to be slower in Group A than in Group B. We can tabulate, for each block of trials in the group situation, the average difference between the median judgments of the three members. In the case of any given block of trials, is the average difference larger for Group A than for Group B?

Experiment XXX

Problem Solving in a Group Situation[9]

Purpose. To compare the problem-solving behavior of subjects working alone and in a group situation.

Materials. Two sets of problems (say, three problems in each set) are needed. The most suitable are those which require a series of successive steps for solution. Such problems easily lend themselves to group discussion since each step is considered and taken up separately by the group. A series of such problems, which were successfully used in this type of experiment, may be found in Shaw's paper, cited in Footnote 9. The experimenter needs prearranged score sheets for recording his observations. On this score sheet should be listed the categories of behavior which he wishes to observe, e.g., suggestions made by each individual, types of criticisms advanced against the suggestions, method of arriving at a decision, etc. If equipment is available, the ideal procedure is to record the discussion of the group on a wire recorder or similar instrument. In conjunction with notes based on visual observations, such records will constitute a full and permanent record of the activities of the group.

Experimental Design. This experiment can be best carried out with the aid of a counterbalanced design. Each subject serves both individually and in a group (groups of four or five are suitable). Half the subjects first work alone and then in a group; for the other half of the subjects,

[9] This experiment is based on a paper by M. E. Shaw, A comparison of individuals and small groups in the rational solution of complex problems, *Amer. J. Psychol.*, 1932, *44*:491-504.

the sequence of conditions is reversed. The two sets of problems—Set A and Set B—should be used also in a counterbalanced order. The total experimental design may be summarized briefly.

Group I: Half the subjects in this group first work individually on Set A and then work as a group on Set B, the other half of the subjects first work individually on Set B and then as a group on Set A.

Group II: Half the subjects first work as a group on Set A and then individually on Set B; the other half of the subjects first work as a group on Set B and then individually on Set A.

This design controls both practice effects and possible differences in the ease of solving the two sets of problems. For laboratory purposes, the design, of course, may be simplified considerably. We may, for example, simply divide the members of the class into group workers and individual workers. Half of the individual workers solve Set A, half solve Set B. Similarly, half the groups work on Set A, the other half work on Set B. In this simplified design, the factor of individual differences in problem-solving ability is not controlled and the experimenter must hope that such differences will be random and not affect the results of the experiment systematically.

Procedure: Individual Work. The subject is handed the problems with the proper instructions. He is urged to work as quickly and as accurately as he can, to record his answers and also to indicate the method he used in solving the problem. The subject may be urged to "think aloud" so that the experimenter may follow his methods of solution step by step. The time needed for solution is recorded. A time limit should be imposed and the problem considered failed if not solved during that time.

Procedure: Group Work. The members of the group are seated around a table and are given the following instructions:

"This is an experiment in problem solving. The materials for the problems are laid out on this table. We want you to work on the solutions as a coöperative group. Try to work as quickly, but also as accurately as you can. Each of the problems has a solution. One of you will serve as chairman to manipulate the materials and guide the group discussion. Each member of the group should contribute to the best of his ability to the solution of the problem."

The activities of the group should be observed either by the experimenter or a "note-taker." The note-taker should do his work inconspicuously and in no way participate in the group discussion. (In experiments of this kind, a "one-way vision" screen is most useful; the experimenter

may watch the group while the members are not aware that they are under observation.) The time taken to solve each problem is noted and again a time limit is imposed. In both individual and group situations, the time limits should be generous so as to afford full opportunity for various methods of solution to be tried out.

Treatment of Results. The two basic sets of data are: (1) the number of problems solved by subjects working alone and the number of problems solved in the group situation, (2) the average time required to solve a problem in the two situations. Group and individual work can be compared with respect to these two scores and the significance of the differences tested.

The analysis of the qualitative data is, of course, a challenge to the ingenuity of the experimenter. Some questions, however, immediately suggest themselves:

1. At what stage in the solution of the problems did individual workers and coöperating groups commit errors? Did the group serve the function of staving off mistakes?
2. How many correct and how many incorrect suggestions were made in the group discussion? What percentage of each was accepted and what percentage rejected? What percentage of rejections originated with the author of the suggestion, and what percentage with other members of the group? In other words, does the group serve to facilitate self-criticism as well as provide critical evaluation by others?
3. What role did the chairman play in the discussion? Did he serve merely as a moderator or did he take an active part in guiding the discussion? Did the leadership shift to a member with greater skill in the manipulation of the material? Did the group divide into active and passive members?

If data on several groups are available, it is interesting to relate their productivity to these and similar characteristics of the group activity.

We have listed only a few ways in which the quantitative comparison of individual and group performance can be supplemented by qualitative observations. Further categories of analysis will undoubtedly suggest themselves in the course of the experiment.

General References

Bird, C., *Social psychology*, New York: D. Appleton-Century Co., 1940.

Britt, S. H., *Social psychology of modern life*, New York: Farrar and Rinehart, 1941.

Murphy, G., Murphy, L. B., and Newcomb, T. M., *Experimental social psychology*, New York: Harper & Brothers, 1937.

Formation of Norms

Schonbar, R. A., The modification of judgments in a group situation, *J. Exper. Psychol.*, 1947, *37*:69-80.

Sherif, M., A study of some social factors in perception, *Arch. Psychol.*, 1935, No. 187.

Sherif, M., *The psychology of social norms*, New York: Harper & Brothers. 1936.

Suggestion and Suggestibility

Aveling, F., and Hargreaves, H. L., Suggestibility with and without prestige in children, *Brit. J. Psychol.*, 1921, *12*:53-75.

Bird, C., *Social psychology*, New York: D. Appleton-Century Co., 1940, Chap. 3.

Brown, W., Individual and sex differences in suggestibility, *Univ. Calif. Publ. Psychol.*, 1916, *2*:291-430.

Coffin, T. E., Some conditions of suggestion and suggestibility, *Psychol. Monogr.*, 1941, *53*, No. 241.

Hull, C. L., *Hypnosis and suggestibility*, New York: D. Appleton-Century Co., 1933.

Hurlock, E. B., The suggestibility of children, *J. Genet. Psychol.*, 1930, *37*:59-74.

Krech, D., and Crutchfield, R. S., *Theory and problems of social psychology*, New York: McGraw-Hill Book Co., 1948, Chap. 9.

Marple, C. H., The comparative susceptibility of three age levels to the suggestion of group versus expert opinion, *J. Soc. Psychol.*, 1933, *4*:176-186.

Murphy, G., Murphy, L. B., and Newcomb, T. M., *Experimental social psychology*, New York: Harper & Brothers, 1937, Chap. 4.

Saadi, M., and Farnsworth, P. R., The degrees of acceptance of dogmatic statements and preferences for their supposed makers, *J. Abn. Soc. Psychol.*, 1934, *29*:143-150.

Sherif, M., An experimental study of stereotypes, *J. Abn. Soc. Psychol.*, 1935, *29*:371-375.

Work in a Group Situation

General Reference

Dashiell, J. F., Experimental studies of the influence of social situations on the behavior of individual human adults. In C. Murchison (ed.),

A handbook of general experimental psychology, Worcester: Clark University Press, 1934.

Audience Effects

Gates, G. S., The effect of an audience upon performance, *J. Abn. Soc. Psychol.*, 1924, *18*:334-342.

Laird, D. A., Changes in motor control and individual variations under the influence of 'razzing,' *J. Exper. Psychol.*, 1923, *6*:236-246.

Pessin, J., and Husband, R. W., Effects of social stimulation on human maze learning, *J. Abn. Soc. Psychol.*, 1933, *28*:148-154.

Travis, L. E., The effect of a small audience upon eye-hand coördination, *J. Abn. Soc. Psychol.*, 1925, *20*:142-146.

Presence of Co-workers

Allport, F. H., The influence of the group upon association and thought, *J. Exper. Psychol.*, 1920, *3*:159-182.

Dashiell, J. F., An experimental analysis of some group effects, *J. Abn. Soc. Psychol.*, 1930, *25*:190-199.

Competitive Work Situations

Hurlock, E. B., The use of rivalry as an incentive, *J. Abn. Soc. Psychol.*, 1927, *22*:278-290.

Maller, J. B., Coöperation and competition, *Teach. Coll. Contr. Educ.*, 1929, No. 384.

Sims, V. M., The relative influence of two types of motivation on improvement, *J. Educ. Psychol.*, 1928, *19*:480-484.

Whittemore, I. C., The influence of competition on performance, *J. Abn. Soc. Psychol.*, 1924, *19*:236-253.

Coöperative Work Situations

Shaw, M. E., A comparison of individuals and small groups in the rational solution of complex problems, *Amer. J. Psychol.*, 1932, *44*:491-504.

South, E. B., Some psychological aspects of committee work, *J. Appl. Psychol.*, 1927, *11*:348-368; 437-464.

Watson, G. B., Do groups think more efficiently than individuals? *J. Abn. Soc. Psychol.*, 1928, *23*:328-336.

Name Index

Subject Index